PCT/GB 2003 / 0 0 1 5 8 9

10 Rec'd PCT/PTO 17 OCT 2004

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

RECD 16 MAY 2003
PCT

PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

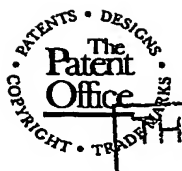
Signed Stephen Hurdle

Dated 29 April 2003

BEST AVAILABLE COPY

Patents Form 1/77

Patents Act 1977
(Rule 16)



THE PATENT OFFICE
B
11 APR 2002
NEWPORT

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

The Patent Office

Cardiff Road
Newport
South Wales
NP10 8QQ

1. Your reference P0150-GB01

2. Patent application number *(The Pat* 0208332.7 — 11 APR 2002

3. Full name, address and postcode of the or of each applicant *(underline all surnames)*

   OXFORD GLYCOSCIENCES (UK) LTD
   THE FORUM
   86 MILTON PARK
   ABINGDON, OXON OX14 4RY
   UNITED KINGDOM

   Patents ADP number *(if you know it)* ~~7112386002~~ 7112386002

   If the applicant is a corporate body, give the country/state of its incorporation: ENGLAND AND WALES

4. Title of the invention

   PROTEIN

5. Name of your agent *(if you have one)*

   "Address for service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*

   BLAKEY, ALISON JANE
   OXFORD GLYCOSCIENCES (UK) LTD
   THE FORUM
   86 MILTON PARK
   ABINGDON, OXON OX14 4RY
   UNITED KINGDOM

   Patents ADP number *(if you know it)* ~~7112386002~~ 7081219002

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and *(if you know it)* the or each application number

Country	Priority application number *(if you know it)*	Date of filing *(day / month / year)*

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application	Date of filing *(day / month / year)*

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer 'Yes' if:*
   a) any applicant named in part 3 is not an inventor, or
   b) there is an inventor who is not named as an applicant, or
   c) any named applicant is a corporate body.
   See note (d))

   YES

Patents Form 1/77

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form	-
Description	41
Claim(s)	3
Abstract	1
Drawing(s)	8 + 8

10. If you are also filing any of the following, state how many against each item.

Priority documents	
Translations of priority documents	
Statement of inventorship and right to grant of a patent *(Patents Form 7/77)*	
Request for preliminary examination and search *(Patents Form 9/77)*	
Request for substantive examination *(Patents Form 10/77)*	
Any other documents *(please specify)*	

11. I/We request the grant of a patent on the basis of this application.

Signature Alison Bailey Date 10/4/02

12. Name and daytime telephone number of person to contact in the United Kingdom MARY GADSDEN 01235 208127

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

a) *If you need help to fill in this form or you have any questions, please contact the Patent Office on 08459 500505.*

b) *Write your answers in capital letters using black ink or you may type them.*

c) *If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.*

d) *If you have answered 'Yes' Patents Form 7/77 will need to be filed.*

e) *Once you have filled in the form you must remember to sign and date it.*

f) *For details of the fee and ways to pay please contact the Patent Office.*

PROTEIN

The present invention relates to the use of a protein (NKCC1) identified in breast cancer membrane preparations and pancreatic cancer membrane preparations, compositions comprising the protein, including vaccines and antibodies that are immunospecific for the protein. The use of the protein in the diagnosis, screening, treatment and prophylaxis of breast and/or pancreatic cancer is also provided.

Breast Cancer

Breast cancer is the second leading cause of cancer death for women in the U.S; approximately 40,000 women in the US die from the disease each year. Breast cancer is the leading cause of cancer death for US women between the ages of 20 and 59, and the leading cause of cancer death for women world-wide. In the United Kingdom there are over 38,000 new cases each year accounting for more than one in four of all cancer cases in women. The lifetime risk of breast cancer in women in the United Kingdom is 1 in 9 (1 in 8 in the USA). Ninety-nine percent of breast cancers occur in women with the risk of development increasing with age. The annual cost of breast cancer treatment in the United States is approximately $10 billion (Fuqua, et. al. 2000, American Association for Cancer Research, www.aacr.org, USA). Breast cancer incidence has been rising over the past five decades, but recently it has reached a plateau. This may reflect a period of earlier detection of breast cancers by mammography. Mammography screening does not prevent or cure breast cancer; however, it may detect the disease before symptoms occur. Breast cancer tumours can exist for six to ten years before they grow large enough to be detected by mammography. In addition, mammography is less effective in younger women than in older women. Studies show that regular mammography screening of women between the ages of 50 and 69 reduces breast cancer mortality by approximately 30% in this age group. However, women with breast cancer will not benefit from mammography screening unless they have access to appropriate treatment. All women are at risk for breast cancer and approximately 90% of women who develop breast cancer do not have a family history of the disease. Factors that increase a woman's risk of breast cancer include: older age, earlier age at menarche, later age at menopause, nulliparity, later age at first full-term pregnancy, daily alcohol consumption, long-term use of hormonal replacement therapy, postmenopausal obesity, infrequent physical activity, ionising radiation, genetic factors and family history of breast or ovarian cancer. Although scientists have discovered some risk factors for breast cancer, most factors account for only small increases in a person's chances of developing breast cancer for which there is no cure.

Surgery, radiation therapy, hormone therapy, and chemotherapy are the most common treatments for breast cancer. When chemotherapy is given after surgery (adjuvant therapy) it can reduce the chance of cancer recurrence. Chemotherapy can also be used as the main treatment for women whose cancer is widespread when it is found, or spreads widely after initial treatment. Neoadjuvant chemotherapy is given before surgery, often to shrink the tumour and make it easier

to remove. It is often more effective to use several drugs, rather than a single drug alone. The most commonly used combinations are: cyclophosphamide, methotrexate, and fluorouracil (CMF) cyclophosphamide, doxorubicin (Adriamycin™), and fluorouracil (CAF) doxorubicin (Adriamycin) and cyclophosphamide (AC), with or without paclitaxel (Taxol™) doxorubicin (Adriamycin), followed by CMF.

The hormone oestrogen can increase the growth of breast cancer cells in some women. A drug such as tamoxifen, which blocks the effect of oestrogen, is given to counter this growth. Another newer drug, raloxifene, also blocks the effect of oestrogen on breast tissue and breast cancer. There is increasing evidence that these anti-oestrogen treatments may also have a role in chemoprevention of breast cancer in high-risk individuals.

Trastuzumab (Herceptin™) is a new immunotherapeutic agent that attaches to a growth factor receptor known as c-erbB2/HER2/neu, which is present in small amounts on the surface of normal breast cells and at much higher levels in some breast cancers. This protein can cause the cancer to grow and spread faster. Herceptin can stop the c-erbB2/HER2/neu protein from promoting breast cancer cell growth. It may also help the immune system to better attack the cancer. Herceptin is currently started after standard hormonal or chemotherapy is no longer working (American Cancer Society, 2000, USA, www.cancer.org).

Pancreatic Cancer

Pancreatic cancer is currently the 5th highest cause of cancer deaths in Europe and USA. In the USA it occurs with a frequency of 9.2 cases per 100,000 population and accounts for approx. 25,000 deaths per year.

The pancreas can be divided into exocrine and endocrine regions with the exocrine region making up the majority of the pancreas with a ratio of 99:12 exocrine to endocrine cells. The exocrine region produces enzyme rich fluid or juice that is secreted into the duodenum to aid digestion. The endocrine region consists of the Islets of Langerhans and produces hormones such as insulin, glucagon and somatostatin that regulate the use of sugars and starches within the body. Approximately 95% of pancreatic cancers are exocrine (90% of which are duct cell carcinomas) and are histologically classified as adenocarcinomas.

Exocrine pancreatic cancer is not usually cured by standard therapy, and the present therapies available have many non-desirable side-effects.
In contrast to exocrine cancers, endocrine cancers (islet cell carcinomas) are usually slower growing and are often treatable. However, endocrine cancers are rare accounting for fewer than 1,000 cases per year in the USA.

Given its incidence and almost universal fatality, substantially increased research efforts are clearly warranted to understand, prevent, and control this disease.

Causes of Pancreatic Cancer

Pancreatic cancer is a rapidly fatal disease with the median age at diagnosis being 71 years, and its incidence varies by race, gender, and geography. The disease occurs more often in

African Americans than in whites and in men more than in women; incidence rates around the world vary approximately 30-fold. African Americans have the highest pancreatic cancer rate in the world.

In addition to ageing, there are four probable risk factors for pancreatic cancer: family history, cigarette smoking, long-standing diabetes, and hereditary and chronic pancreatitis:

Family history - People in affected families have about a three-fold higher risk compared with the general population.

Cigarette smoking - Smoking is believed to cause about one quarter to one third of pancreatic cancers. People who smoke for twenty years or more have double the risk of those who have never smoked, and recent evidence indicates that this risk may be even higher when certain genetic polymorphisms are present.

Long-standing diabetes - There is a two-fold increased risk of pancreatic cancer among people who were diagnosed with diabetes mellitus at least five years before their diagnosis of pancreatic cancer. This observation suggests that diabetes may be an independent risk factor for pancreatic cancer, as well as a possible consequence of the disease. The mechanism involved, however, is unclear.

Chronic pancreatitis - Pancreatic cancer risk among individuals with hereditary pancreatitis or non-hereditary chronic pancreatitis is about 50 times and 16 to 20 times higher, respectively, than those without chronic pancreatitis.

Studies also have implicated a number of other factors, including diet and nutrition, heavy alcohol consumption, and certain occupational exposures, but these findings have been inconsistent.

Diagnosis

Diagnosis of pancreatic cancer is currently difficult as early symptoms are similar to those of other disorders including chronic pancreatitis, hepatitis, gallstones and diabetes mellitus. Often, by the time a correct diagnosis has been made, the cancer has spread to the lymph nodes and the liver. Diagnosis is currently made after a series of tests including ultrasonography, fine-needle biopsy, CT scans, MRI scans, endoscopic retrograde cholangiopancreatography (ECRP), and percutaneous transhepatic cholangiography. In particular, several potential pancreatic cancer biomarkers have been identified such as: carcino-embryonic antigen (CEA), CA 19-9, pancreatic oncofoetal antigen (POA), alphafetoprotein (AFP), neuron-specific enolase (NSE), and CA-195. However, known biological markers are less sensitive than imaging techniques, and they lack specificity. Better screening and diagnostic techniques are urgently needed. There is however, some evidence to suggest that pancreatic cancer may progress from flat ductal lesions, to papillary ductal lesions without atypia, to papillary ductal lesions with atypia, and finally, to infiltrating adenocarcinoma. Importantly, the existence of such a progression suggests that the

ability to detect a curable precursor lesion and early cancers with a molecular test may be possible, perhaps by mutant k-ras shed from the pancreatic intraepithelial neoplasia and detected in stool, duodenal fluid, or pancreatic juice samples (Wilentz, R. and Hruban, R.1998, Surg. Oncol. Clin. N. Am. 7: 43-65).

Disease Staging and Prognosis

Staging is the process of finding out how far the cancer has spread. The staging system of the American Joint Committee on Cancer (AJCC), also known as the TNM system, is the one used most often for pancreatic cancer. The TNM system for staging gives three key pieces of information:

The letter T stands for tumour and is usually followed by a number from 0 to 4. A higher number relates to a larger tumour size and how far it has spread within the pancreas and to nearby organs.

The letter N stands for spread to lymph nodes and when followed by a 0 or 1 indicates no spread to nearby or more distant nodes, respectively.

The letter M is for metastasis and when followed by a 0 or 1, shows that the cancer has not spread to distant lymph nodes (other than those near the pancreas), or has spread to distant nodes and organs, respectively.

To make this information somewhat clearer, the TNM descriptions can be grouped together into a simpler set of stages, labelled stage 0 through stage IV (0-4). In general, the lower the number, the less the cancer has spread. A higher number, such as stage IV (4), means a more serious cancer. (American Cancer Society, 2000, USA, www.cancer.org)

Pancreatic Cancer Survival Rates

Typically, diagnosis of pancreatic cancer is at a later stage of the disease due to the non-specific symptoms associated with the disease. Indeed, the prognosis for patients is generally poor as it is a this disease which is characterised by metastatic tumour spread even when the primary tumour is very small. The five-year survival rates for all stages of pancreatic cancer are 3.6%; for localised tumours, 12%; regional spreading, 4.8% and 1.6% for tumours that have metastasised to distant lymph nodes and organs (www.sciam.com).

There are a number of variants of pancreatic adenocarcinoma, including adenosquamous carcinoma, acinar cell carcinoma, and giant cell carcinoma. These variants account for only a small percentage of pancreatic ductal carcinomas. Patients diagnosed with adenosquamous and giant cell carcinomas tend to have poorer prognoses, while acinar cell cancer is associated with a better prognosis than is typical ductal adenocarcinoma.

At the genetic level, a large number of resected pancreatic adenocarcinomas have been examined for genetic alterations in cancer-causing genes. In general, these genes can be divided into three broad groups (Hruban R *et al.*, 1998 Surg. Oncol. Clin. N. Am. 7:1-23): tumor suppressors, oncogenes and DNA-mismatch repair genes The tumor suppressor genes p53, p16, and DPC4 are inactivated frequently in sporadic carcinoma of the pancreas. p53 is a well-

characterized tumor suppressor gene whose function appears to be inactivated in up to 75% of all pancreatic carcinomas. Recently Hahn *et al* (Science 1996,;271: 350-353) identified a new tumor suppressor gene, designated DPC4, which appears to be more specific than p53 and p16 for pancreatic cancer. DPC4 resides on chromosome 18q, a chromosome shown by allelotyping to be lost in 90% of pancreatic carcinomas.

Activating point mutations in the k-ras oncogene are the most common genetic alteration identified in pancreatic cancer and have been found in 80% to 100% of pancreatic cancers, most of these being mutations in codon 12.

A recent report noted that 4% of pancreatic adenocarcinomas were characterized by disorders of DNA repair genes (Goggins, M. et al., 1998, Am J Pathol. 152: 1501-1507). Limited data suggest that this subgroup of tumors may have a more favorable prognosis than that of the typical adenocarcinoma without mutation in DNA repair genes.

Treatment

The treatment options available are varied and depend on the stage or progression of the cancer.

Resection: Pancreaticoduodenectomy or "Whipple" procedure – this involves surgical removal of the head and neck of the pancreas, the duodenum and adjacent structures.
Total pancreatectomy – this involves surgical removal of the entire pancreas, duodenum, common bile duct, gall bladder, spleen and most of the adjacent lymph nodes
Regional pancreatectomy – this involves surgical removal of the pancreas, lymph nodes, some veins and arteries supplying the area. This procedure is particularly effective for small tumours at the head of the pancreas.

Radiotherapy: The use of high energy x-rays is common in an attempt to control the disease, provide relief from symptoms and prolong survival. In particular, the use of intraoperative radiation therapy (IORT) has been shown to prolong the survival time of the patients. In this method radiation is used during surgery maximising the effect of the radiation on the tumour whilst avoiding damage to nearby tissue.

Chemotherapy: Chemotherapy is usually used in advanced pancreatic cancer. The combination of 5-fluorouracil (5-FU) and intraoperative/standard radiation therapy is able to prolong survival. Gemcitibine is currently one of the most effective chemotherapy drugs available for pancreatic cancer. Some success has been found with combination therapies such as gemcitibine-auristatin-PE combinations however, no drug regime has yet shown a consistently good result in clinical studies.

Biological therapy: More recently, attention has focused on the development of immunological therapies that aim to enhance the patients immune system enabling the immune system to fight the cancer, suppress body responses which allow the cancer growth and make the cancerous cells more vulnerable to destruction by the immune system.

Therapeutic monoclonal antibodies have received particular attention in recent times. For example, a combination treatment is currently in clinical trials. This involves a combination of a monoclonal antibody, IMC-C225, directed against the epidermal growth factor (EGF) receptor (the EGF receptor is over-expressed in pancreatic cancer tumours), and 5-fluorouracil (5-FU), setting a foundation for future therapeutic application of the IMC-C225 agent in the treatment of pancreatic carcinoma. Indeed not only the latter, but also a number of other agents are presently undergoing clinical trials (http://www.pancreatica.org). It has also been reported that the HER-2/neu oncogene was over-expressed in 21% of 154 patients with pancreatic adenocarcinoma (Safran, H. et al., 2001 Amer. J. Clin. Oncology 24: 496-499) where it has been suggested that therapeutic evaluation of recombinant humanized anti-HER2 antibody (Herceptin) with such patients may be warranted. In addition, studies with biological response modifiers are ongoing including interferon alpha, recombinant interleukin-2 (and combinations of these agents with 5-FU and cyclophosphamide), leucovorin and granulocyte macrophage-colony stimulating factor. Thus, the identification of other suitable targets or antigens for immunotherapy of pancreatic cancer has become increasingly important.

Hormone therapy: Both estrogens and androgens may affect the growth of pancreatic tissue. There have been some reports of occasional success in treating pancreatic cancer with tamoxifen, a hormonal therapy drug commonly used for women with breast cancer. There hasn't been any overall benefit for most people, however. Other hormones that are produced by the stomach and intestines may also affect the growth of pancreatic cancer. Treatments designed to block these hormones are currently being studied (see both www.cancer.org and Andren-Sandberg A. & Backman PL: Hormonal therapy and immunotherapy, in Howard JM, Idezuki Y, Ihse I, Prinz RA (eds): Surgical Diseases of the Pancreas ed 3. Baltimore, Williams and Wilkins, 1998, pp. 613-622).

Therapeutic Challenges

While survival has improved for patients with most other gastrointestinal cancers, the five-year survival of patients with pancreatic adenocarcinoma remains less than five percent. Nominal therapeutic advances have been made in recent years. A new drug, gemcitabine, has improved the quality of life and modestly affected the survival of patients with locally advanced or metastatic pancreatic cancer. Despite these approaches, however, the overall impact of therapy for pancreatic cancer is quite limited.

Basic research efforts over the past few years have shown that pancreatic adenocarcinomas, like other major human neoplasms, result from accumulating genetic lesions that lead to tumour development and promote progression. Though a number of these germline and somatic tumour-associated alterations have been identified, significant gaps exist in our understanding of how these alterations initiate the process of pancreatic carcinogenesis. In particular, it remains a challenge to better understand and determine how the genetics and molecular biology of pancreatic cancer can be harnessed for therapeutic gain. Given the highly aggressive clinical characteristics and lack of effective therapies for pancreatic cancer, advancing

our knowledge in these areas is of special urgency.

The major challenges in breast and pancreatic cancer treatments are to improve early detection rates, to find new non-invasive markers that can be used to follow disease progression and identify relapse, and to find improved and less toxic therapies, especially for more advanced disease where 5 year survival is still very poor. There is a great need to identify targets which are more specific to the cancer cells, ideally ones which are expressed on the surface of the tumour cells so that they can be attacked by promising new approaches like immunotherapeutics and targeted toxins.

An ideal protein target for cancer immunotherapy should have a restricted expression profile in normal tissues and be over-expressed in tumours, such that the immune response will be targeted to tumour cells and not against other organs. In addition, the protein target should be exposed on the cell surface, where it will be accessible to therapeutic Tumour specific proteins have been identified for a number of cancer types, by using techniques such as differential screening of cDNA (Hubert, R.S., *et al.*, 1999, *Proc. Natl. Acad, Sci. USA* **96**, 14523-14528; Lucas, S., *et al. Int. J. Cancer* **87**, 55-60 (2000)), and the purification of cell-surface proteins that are recognised by tumour-specific antibodies (Catimel, B., *et al. J. Biol. Chem.* **271**, 25664-25670 (1996)). As an alternative approach to identifying breast cancer and pancreatic cancer antigens, we have used proteomics to characterise the complement of proteins in cell membranes isolated from the breast cancer and pancreatic cancer samples.

NKCC1

NKCC1 is a bumetanide sensitive sodium-potassium-chloride (Na-K-Cl) cotransporter (Payne *et al*, (1995), J Biol Chem, 270(30): pp17977-17985). Two isoforms of the Na-K-Cl cotransporter have been identified; one located on the apical membrane of absorptive epithelia (NKCC2) and one located on the basolateral membrane of secretory epithelia (NKCC1). The function of the Na-K-Cl cotransporters is to provide electroneutral transport of chloride ions across epithelia (in a ratio of 1Na:1K:2Cl), they work in combination with the sodium and potassium channels and the sodium pump to cause a net transport of sodium chloride across membranes. Dysregulation of this transport mechanism can result in diseases such as Cystic Fibrosis and secretory diarrhoea.

The human NKCC1 protein was identified by the Payne *et al* in 1995 (see above) by the screening of a human colon carcinoma cell line (T84). It shows significant similarity to the NKCC1 proteins identified from elasmobranch, mouse, rat, rabbit and flounder.

Both isoforms of the Na-K-Cl cotransporter show a similar structure with large N and C terminal domains and 12 transmembrane segments. NKCC1 is significantly larger than NKCC2 with an extra 80 amino acids at the N terminus, interestingly it is the N terminus that shows the greatest variation between species indicating that this section may not be directly involved in the ion transportation. A recent study has indicated that NKCC1 may have 2 splice variants which show differential expression across a variety of tissue types indicating that differential splicing may play a regulatory role in NKCC1 activity (Vibat et al Anal Biochem 2001;298(2):218-30).

Overexpression of NKCC1 has been reported in asthmatic subjects by gene expression profiling (Dolganov GM *et al*, Genome Res (2001); 11(9): 1473-83). This data was then confirmed using immunohistochemistry, which demonstrated that in asthmatic subjects NKCC1 expression is increased with restricted localisation to goblet cells implicating NKCC1 in the pathophysiology of mucus hypersecretion in asthma.

However, the prior art does not show a cancer-associated alteration of the NKCC1 protein and therefore, does not show the usefulness of NKCC1 in a therapeutic treatment approach to breast cancer and or pancreatic cancer, more specifically to tumour cells derived from epithelial cells typically found in the lining of body organs for example, but not limited to, breast and pancreas.

The present invention is based on the finding that the protein, designated NKCC1, (the sequence of which can be found under the accession number P55011 in the SwissProt database, (available at http://www.expasy.org/), and which is shown in Figure 1) is expressed in breast cancer cell line membranes. In addition, its mRNA shows restricted expression to a few tissues, with elevated expression in breast cancer and pancreatic cancer samples, suggesting that it may be a suitable target for breast and/or pancreatic cancer therapy and diagnosis.

The sequence shown in Figure 1 represents the amino acid and nucleic acid sequences of NKCC1 identified in breast and pancreatic cancer membranes.

Thus, one aspect of the present invention provides a method of screening for and/or diagnosis of breast and/ or pancreatic cancer in a subject and/or monitoring the effectiveness of breast and/ or pancreatic cancer therapy, which method comprises the step of detecting and/or quantifying the amount of a polypeptide in a biological sample obtained from said subject, wherein the polypeptide:

a). comprises or consists of the amino acid sequence shown in Figure 1 (SEQ ID NO:1);

b). is a derivative having one or more amino acid substitutions, modifications, deletions or insertions relative to the amino acid sequence shown in Figure 1 (SEQ ID NO:1); or

c). is a fragment of a polypeptide as defined in a) or b) above, which is at least ten amino acids long;

In another aspect, the present invention provides a method for the prophylaxis and/or treatment of breast cancer and/or pancreatic cancer in a subject, which comprises administering to said subject a therapeutically effective amount of at least one polypeptide as defined a)-c) above.

In a yet another aspect, the present invention provides the use of at least one polypeptide as defined in a)-c) above in the preparation of a composition for use in the prophylaxis and/or treatment of breast cancer and/or pancreatic cancer.

The subject may be a mammal and is preferably a human, although monkeys, apes, cats, dogs, cows, horses and rabbits are within the scope of the present invention.

In the aspects *supra*, the polypeptides, derivatives or fragments thereof may be provided in isolated or recombinant form, and may be fused to other moieties. In particular, fusions of the

polypeptides, derivatives or fragments thereof with localisation-reporter proteins such as the Green Fluorescent Protein (U.S. Patent Nos. 5,625,048, 5,777,079, 6,054,321 and 5,804,387) or the DsRed fluorescent protein (Matz, M. V., *et al.* Nature Biotech. 17:969–973) are specifically contemplated. The polypeptides, derivatives or fragments thereof may be provided in substantially pure form, that is to say free, to a substantial extent, from other proteins. Thus, a polypeptide for use in the present invention may be provided in a composition in which it is the predominant component present (i.e. it is present at a level of at least 50%; preferably at least 75%, at least 80%, at least 85%, at least 90%, or at least 95%; when determined on a weight/weight basis excluding solvents or carriers).

In order to more fully appreciate the present invention, polypeptides within the scope of a)-c) above will now be discussed in greater detail. It will be apparent to one skilled in the art that peptides for use in the invention include NKCC1, and derivatives, fragments and modified forms thereof.

Polypeptides within the scope of a)

A polypeptide within the scope of a), may consist of the particular amino acid sequence given in Figure 1 or may have an additional N-terminal and/or an additional C-terminal amino acid sequence relative to the sequence given in Figure 1.

Additional N-terminal or C-terminal sequences may be provided for various reasons. Techniques for providing such additional sequences are well known in the art. Additional sequences may be provided in order to alter the characteristics of a particular polypeptide. This can be useful in improving expression or regulation of expression in particular expression systems. For example, an additional sequence may provide some protection against proteolytic cleavage. This has been done for the hormone Somatostatin by fusing it at its N-terminus to part of the β galactosidase enzyme (Itakwa *et al.*, (1977) *Science* **198**: 105-63).

Additional sequences can also be useful in altering the properties of a polypeptide to aid in identification or purification. For example, a fusion protein may be provided in which a polypeptide is linked to a moiety capable of being isolated by affinity chromatography. The moiety may be an antigen or an epitope and the affinity column may comprise immobilised antibodies or immobilised antibody fragments which bind to said antigen or epitope (desirably with a high degree of specificity). The fusion protein can usually be eluted from the column by addition of an appropriate buffer.

Additional N-terminal or C-terminal sequences may, however, be present simply as a result of a particular technique used to obtain a polypeptide and need not provide any particular advantageous characteristic to the polypeptide. Such polypeptides are within the scope of the present invention.

Whatever additional N-terminal or C-terminal sequence is present, it is preferred that the resultant polypeptide should exhibit the immunological or biological activity of the polypeptide having the amino acid sequence shown in Figure 1.

Polypeptides within the scope of b)

Turning now to the polypeptides defined in b) above, it will be appreciated by the person skilled in the art that these polypeptides are derivatives of the polypeptide given in a) above, provided that such derivatives preferably exhibit the immunological or biological activity of the polypeptide having the amino acid sequence shown in Figure 1. Alternatively, the biological activity of the polypeptide may be altered. As such, it will be appreciated by one skilled in the art that derivatives can include post-translational modifications, for example but without limitation, phosphorylation, glycosylation and farnesylation.

As used herein, the term "derivative" refers to a polypeptide that comprises an amino acid sequence of a parent polypeptide that has been altered by the introduction of amino acid residue substitutions, deletions or additions, and/or amino acid modifications such as but not limited to, phosphorylation and glycosylation. The derivative polypeptide possesses a similar or identical activity to the parent polypeptide.

As used herein, the term "fragment" refers to a peptide or polypeptide comprising an amino acid sequence of at least 5 amino acid residues (preferably, at least 10 amino acid residues, at least 15 amino acid residues, at least 20 amino acid residues, at least 25 amino acid residues, at least 40 amino acid residues, at least 50 amino acid residues, at least 60 amino residues, at least 70 amino acid residues, at least 80 amino acid residues, at least 90 amino acid residues, at least 100 amino acid residues, at least 125 amino acid residues, at least 150 amino acid residues, at least 175 amino acid residues, at least 200 amino acid residues, at least 250 amino acid residues, at least 300 amino acid residues, at least 350 amino acid residues, at least 400 amino acid residues, at least 450 amino acid residues, at least 500 amino acid residues, at least 550 amino acid residues, at least 600 amino acid residues, at least 650 amino acid residues, at least 700 amino acid residues, at least 750 amino acid residues, at least 800 amino acid residues, at least 850 amino acid residues, at least 900 amino acid residues, at least 950 amino acid residues, at least 1000 amino acid residues, at least 1050 amino acid residues, at least 1100 amino acid residues, at least 1150 amino acid residues or at least 1200 amino acid residues) of the amino acid sequence of a parent polypeptide. Any given fragment of the polypeptide for use in the invention may or may not possess the functional activity of the parent polypeptide.

Alterations in the amino acid sequence of a protein can occur which do not affect the activity of a protein. These include amino acid deletions, insertions and substitutions and can result from alternative splicing and/or the presence of multiple translation start sites and stop sites. Polymorphisms may arise as a result of the infidelity of the translation process. Thus changes in amino acid sequence may be tolerated which do not affect the protein's biological or immunological activity.

The skilled person will appreciate that various changes can often be made to the amino acid sequence of a polypeptide which has a particular activity to produce derivatives (sometimes known as variants or "muteins") having at least a proportion of said activity, and preferably having a substantial proportion of said activity. Such derivatives of the polypeptides described in a) above are within the scope of the present invention and are discussed in greater detail below. They include allelic and non-allelic derivatives.

An example of a derivative of the polypeptide for use in the present invention is a polypeptide as defined in a) above, apart from the substitution of one or more amino acids with one or more other amino acids. The skilled person is aware that various amino acids have similar properties. One or more such amino acids of a polypeptide can often be substituted by one or more other such amino acids without eliminating a desired activity of that polypeptide.

Thus, the amino acids glycine, alanine, valine, leucine and isoleucine can often be substituted for one another (amino acids having aliphatic side chains). Of these possible substitutions, it is preferred that glycine and alanine are used to substitute for one another (since they have relatively short side chains) and that valine, leucine and isoleucine are used to substitute for one another (since they have larger aliphatic side chains which are hydrophobic).

Other amino acids which can often be substituted for one another include:

- phenylalanine, tyrosine and tryptophan (amino acids having aromatic side chains);
- lysine, arginine and histidine (amino acids having basic side chains);
- aspartate and glutamate (amino acids having acidic side chains);
- asparagine and glutamine (amino acids having amide side chains);
- cysteine and methionine (amino acids having sulphur-containing side chains); and
- aspartic acid and glutamic acid can substitute for phospho-serine and phospho-threonine, respectively (amino acids with acidic side chains).

Substitutions of this nature are often referred to as "conservative" or "semi-conservative" amino acid substitutions.

Amino acid deletions or insertions may also be made relative to the amino acid sequence given in a) above. Thus, for example, amino acids which do not have a substantial effect on the biological and/or immunological activity of the polypeptide, or at least which do not eliminate such activity, may be deleted. Such deletions can be advantageous since the overall length and the molecular weight of a polypeptide can be reduced whilst still retaining activity. This can enable the amount of polypeptide required for a particular purpose to be reduced for example, dosage levels can be reduced.

Amino acid insertions relative to the sequence given in a) above can also be made. This may be done to alter the properties of a polypeptide used in the present invention (e.g. to assist in identification, purification or expression, as explained above in relation to fusion proteins).

Amino acid changes relative to the sequence given in a) above can be made using any suitable technique e.g. by using site-directed mutagenesis (Hutchinson et al., 1978, J. Biol. Chem. 253:6551).

It should be appreciated that amino acid substitutions or insertions to the polypeptide for use in the present invention can be made using naturally occurring or non-naturally occurring amino acids. Whether or not natural or synthetic amino acids are used, it is preferred that only L- amino acids are present.

Whatever amino acid changes are made (whether by means of substitution, modification, insertion or deletion), preferred polypeptides for use in the present invention have at least 50% sequence identity with a polypeptide as defined in a) above, more preferably the degree of

sequence identity is at least 75%, at least 80%, at least 85%. Sequence identities of at least 90% or at least 95% are most preferred.

The term identity can be used to describe the similarity between two polypeptide sequences. The degree of amino acid sequence identity can be calculated using a program such as "bestfit" (Smith and Waterman, Advances in Applied Mathematics, 482-489 (1981)) to find the best segment of similarity between any two sequences. The alignment is based on maximising the score achieved using a matrix of amino acid similarities, such as that described by Schwarz and Dayhof (1979) Atlas of Protein Sequence and Structure, Dayhof, M.O., Ed pp 353-358.

A software package well known in the art for carrying out this procedure is the CLUSTAL program. It compares the amino acid sequences of two polypeptides and finds the optimal alignment by inserting spaces in either sequence as appropriate. The amino acid identity or similarity (identity plus conservation of amino acid type) for an optimal alignment can also be calculated using a software package such as BLASTX. This program aligns the largest stretch of similar sequence and assigns a value to the fit. For any one pattern comparison, several regions of similarity may be found, each having a different score. One skilled in the art will appreciate that two polypeptides of different lengths may be compared over the entire length of the longer fragment. Alternatively small regions may be compared. Normally sequences of the same length are compared for a useful comparison to be made.

Where high degrees of sequence identity are present there will be relatively few differences in amino acid sequence. Thus for example they may be less than 20, less than 10, or even less than 5 differences.

Polypeptides within the scope of c)

As discussed *supra*, it is often advantageous to reduce the length of a polypeptide, preferably the resultant reduced length polypeptide still has a desired activity or can give rise to useful antibodies. Feature c) therefore covers fragments of polypeptides a) or b) above for use in the present invention.

The skilled person can determine whether or not a particular fragment has activity. Fragments are at least 10 amino acids long, preferred fragments may be at least 20, at least 50 or at least 100 amino acids long.

A polypeptide as defined herein may be useful as antigenic material, and may be used in the production of vaccines for treatment or prophylaxis of breast cancer and/or pancreatic cancer. Such material can be "antigenic" and/or "immunogenic". Generally, "antigenic" is taken to mean that the protein is capable of being used to raise antibodies or indeed is capable of inducing an antibody response in a subject. "Immunogenic" is taken to mean that the protein is capable of eliciting a protective immune response in a subject. Thus, in the latter case, the protein may be capable of not only generating an antibody response but, in addition, non-antibody based immune responses.

It is well known that is possible to screen an antigenic protein or polypeptide to identify epitopic regions, i.e. those regions which are responsible for the protein or polypeptide's antigenicity or immunogenicity. Methods well known to the skilled person can be used to test fragments and/or homologues and/or derivatives for antigenicity. Thus, the fragments for use in the present

invention may include one or more such epitopic regions or be sufficiently similar to such regions to retain their antigenic/immunogenic properties. Thus, for fragments for use according to the present invention the degree of identity is perhaps irrelevant, since they may be 100% identical to a particular part of a protein or polypeptide, homologue or derivative as described herein. The key issue may be that the fragment retains the antigenic/immunogenic properties of the protein from which it is derived.

Homologues, derivatives and fragments may possess at least a degree of the antigenicity/immunogenicity of the protein or polypeptide from which they are derived.

In a further aspect, the present invention provides the use of a polypeptide as defined herein in the production of a composition for the treatment or prophylaxis of breast cancer and/or pancreatic cancer, wherein the composition is a vaccine. The vaccine optionally comprises one or more suitable adjuvants. Examples of adjuvants well-known in the art include inorganic gels, such as aluminium hydroxide, and water-in-oil emulsions, such as incomplete Freund's adjuvant. Other useful adjuvants will be well known to the skilled person.

In yet further aspects, the present invention provides:

(a) the use of a polypeptide as defined herein in the preparation of an immunogenic composition, preferably a vaccine;

(b) the use of such an immunogenic composition in inducing an immune response in a subject; and

(c) a method for the treatment or prophylaxis of breast cancer and/or pancreatic cancer in a subject, or of vaccinating a subject against breast cancer and/or pancreatic cancer which comprises the step of administering to the subject an effective amount of a polypeptide as defined herein, preferably as a vaccine.

As will be discussed below, the polypeptides used in the present invention will find use in an immunotherapeutic approach to breast and/or pancreatic cancer. The skilled person will appreciate that for the preparation of one or more such polypeptides, the preferred approach will be based on recombinant DNA techniques.

These nucleic acid molecules encoding the polypeptides or fragments thereof may be used in their own right. Thus, in another aspect, the invention provides a method of screening for and/or diagnosis of breast cancer and/or pancreatic cancer in a subject, which method comprises the step of detecting and/or quantifying the amount of a nucleic acid in a biological sample obtained from said subject, wherein the nucleic acid molecule:

d) comprises or consists of the DNA sequence shown in Figure 1 (SEQ ID NO:2) or its RNA equivalent;

e) a sequence which codes for a derivative or fragment of an amino acid molecule as defined in (i)(a) or (b) (SEQ ID NO:2).

f) a sequence which is complementary to the sequences of d) or e);

g) a sequence which codes for the same polypeptide, as the sequences of d) or e); or

h) a sequence which shows substantial identity with any of those of d), e), f) or g).

The term identity can also be used to describe the similarity between two individual DNA sequences. The 'bestfit' program (Smith and Waterman, Advances in applied Mathematics, 482-489 (1981)) is one example of a type of computer software used to find the best segment of similarity between two nucleic acid sequences, whilst the GAP program enables sequences to be aligned along their whole length and finds the optimal alignment by inserting spaces in either sequence as appropriate. It is preferred if sequences which show substantial identity with any of those of d), e) and f) have e.g. at least 50%, at least 75%, at least 80%, at least 85%, at least 90% or at least 95% sequence identity.

In a further aspect, the present invention provides a method for the prophylaxis and/or treatment of breast cancer and/or pancreatic cancer in a subject, which comprises administering to said subject a therapeutically effective amount of at least one nucleic acid as defined above.

In yet another aspect, the present invention provides the use of at least one nucleic acid as defined above in the preparation of a composition for use in the prophylaxis and/or treatment of breast cancer and/or pancreatic cancer.

These nucleic acid molecules are now discussed in greater detail.

The polypeptides used in the present invention can be coded for by a large variety of nucleic acid molecules, taking into account the well-known degeneracy of the genetic code. All of these molecules can be used in the present invention. They can be inserted into vectors and cloned to provide large amounts of DNA or RNA for further study. Suitable vectors may be introduced into host cells to enable the expression of polypeptides used in the present invention using techniques known to the person skilled in the art.

The term 'RNA equivalent' when used above indicates that a given RNA molecule has a sequence which is complementary to that of a given DNA molecule, allowing for the fact that in RNA 'U' replaces 'T' in the genetic code. The nucleic acid molecule may be in isolated, recombinant or chemically synthetic form.

Techniques for cloning, expressing and purifying proteins and polypeptides are well known to the skilled person. DNA constructs can readily be generated using methods well known in the art. These techniques are disclosed, for example in J. Sambrook *et al*, *Molecular Cloning 2nd Edition*, Cold Spring Harbour Laboratory Press (1989); in Old & Primrose *Principles of Gene Manipulation* 5th Edition, Blackwell Scientific Publications (1994); and in Stryer *Biochemistry* 4th Edition, W H Freeman and Company (1995). Modifications of DNA constructs and the proteins expressed such as the addition of promoters, enhancers, signal sequences, leader sequences, translation start and stop signals and DNA stability controlling regions, or the addition of fusion partners may then be facilitated.

Normally the DNA construct will be inserted into a vector, which may be of phage or plasmid origin. Expression of the protein is achieved by the transformation or transfection of the vector into a host cell, which may be of eukaryotic or prokaryotic origin. Such vectors and suitable host cells form further aspects for use in the present invention.

The nucleotide sequences for use in the present invention, including DNA and RNA, and comprising a sequence encoding a polypeptide as defined herein (or a derivative, fragment, homologue or analogue thereof), may be synthesised using methods known in the art, such as

using conventional chemical approaches or polymerase chain reaction (PCR) amplification. The nucleotide sequences for use in the present invention also permit the identification and cloning of the gene encoding a polypeptide as defined herein from any species, for instance by screening cDNA libraries, genomic libraries or expression libraries.

Knowledge of the nucleic acid structure can be used to raise antibodies and for gene therapy. Techniques for this are well-known by those skilled in the art, as discussed in more detail herein.

By using appropriate expression systems, polypeptides for use in the present invention may be expressed in glycosylated or non-glycosylated form. Non-glycosylated forms can be produced by expression in prokaryotic hosts, such as *E. coli*.

Polypeptides comprising N-terminal methionine may be produced using certain expression systems, whilst in others the mature polypeptide will lack this residue.

Preferred techniques for cloning, expressing and purifying a polypeptide used in the present invention are summarised below:

Polypeptides may be prepared natively or under denaturing conditions and then subsequently refolded. Baculoviral expression vectors include secretory plasmids (such as pACGP67 from Pharmingen), which may have an epitope tag sequence cloned in frame (e.g. myc, V5 or His) to aid detection and allow for subsequent purification of the protein. Mammalian expression vectors may include pCDNA3 and pSecTag (both Invitrogen), and pREP9 and pCEP4 (invitrogen). *E. coli* systems include the pBad series (His tagged - Invitrogen) or pGex series (Pharmacia).

In addition to nucleic acid molecules coding for polypeptides used in the present invention, referred to herein as "coding" nucleic acid molecules, the present invention also includes nucleic acid molecules complementary thereto. Thus, for example, both strands of a double stranded nucleic acid molecule are included in the present invention (whether or not they are associated with one another). Also included are mRNA molecules and complementary DNA Molecules (e.g. cDNA molecules).

The use of nucleic acid molecules which can hybridise to any of the nucleic acid molecules discussed above is also covered by the present invention. Such nucleic acid molecules are referred to herein as "hybridising" nucleic acid molecules. Hybridising nucleic acid molecules can be useful as probes or primers, for example.

Desirably such hybridising molecules are at least 10 nucleotides in length and preferably are at least 25 or at least 50 nucleotides in length. The hybridising nucleic acid molecules preferably hybridise to nucleic acids within the scope of d), e), f), g) or h) above specifically.

Desirably the hybridising molecules will hybridise to such molecules under stringent hybridisation conditions. One example of stringent hybridisation conditions is where attempted hybridisation is carried out at a temperature of from about 35°C to about 65°C using a salt solution that is about 0.9 molar. However, the skilled person will be able to vary such conditions as appropriate in order to take into account variables such as probe length, base composition, type of ions present, etc. For a high degree of selectivity, relatively stringent conditions are used to form the duplexes, such as low salt or high temperature conditions. As used herein, "highly stringent conditions" means hybridisation to filter-bound DNA in 0.5 M $NaHPO_4$, 7% sodium dodecyl

sulphate (SDS), 1 mM EDTA at 65 °C, and washing in 0.1xSSC/0.1% SDS at 68 °C (Ausubel F.M. *et al.*, eds., 1989, Current Protocols in Molecular Biology, Vol. I, Green Publishing Associates, Inc., and John Wiley & Sons, Inc., New York, at p. 2.10.3). For some applications, less stringent conditions for duplex formation are required. As used herein "moderately stringent conditions" means washing in 0.2xSSC/0.1% SDS at 42 °C (Ausubel *et al.*, 1989, *supra*). Hybridisation conditions can also be rendered more stringent by the addition of increasing amounts of formamide, to destabilise the hybrid duplex. Thus, particular hybridisation conditions can be readily manipulated, and will generally be chosen depending on the desired results. In general, convenient hybridisation temperatures in the presence of 50% formamide are: 42 °C for a probe which is 95 to 100% identical to the fragment of a gene encoding a polypeptide as defined herein, 37°C for 90 to 95% identity and 32 °C for 70 to 90% identity. In the preparation of genomic libraries, DNA fragments are generated, some of which will encode parts or the whole of a polypeptide as defined herein. The DNA may be cleaved at specific sites using various restriction enzymes. Alternatively, one may use DNAse in the presence of manganese to fragment the DNA, or the DNA can be physically sheared, as for example, by sonication. The DNA fragments can then be separated according to size by standard techniques, including but not limited to agarose and polyacrylamide gel electrophoresis, column chromatography and sucrose gradient centrifugation. The DNA fragments can then be inserted into suitable vectors, including but not limited to plasmids, cosmids, bacteriophages lambda or T_4, and yeast artificial chromosomes (YACs). (See, for example, Sambrook *et al.*, 1989, *Molecular Cloning*, A Laboratory Manual, 1D Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York; Glover, D.M. (ed.), 1985, *DNA Cloning: A Practical Approach*, MRL Press, Ltd., Oxford, U.K. Vol. I, II; Ausubel F.M. *et al.*, eds., 1989, Current Protocols in Molecular Biology, Vol. I, Green Publishing Associates, Inc., and John Wiley & sons, Inc., New York). The genomic library may be screened by nucleic acid hybridisation to labelled probe (Benton & Davis, 1977, *Science* 196:180; Grunstein & Hogness, 1975, *Proc. Natl. Acad. Sci. U.S.A.* 72:3961).

Manipulation of the DNA encoding a protein is a particularly powerful technique for both modifying proteins and for generating large quantities of protein for purification purposes. This may involve the use of PCR techniques to amplify a desired nucleic acid sequence. Thus the sequence data provided herein can be used to design primers for use in PCR so that a desired sequence can be targeted and then amplified to a high degree.

Typically, primers will be at least five nucleotides long and will generally be at least ten nucleotides long (e.g. fifteen to twenty-five nucleotides long). In some cases, primers of at least thirty or at least thirty-five nucleotides in length may be used.

As a further alternative chemical synthesis may be used. This may be automated. Relatively short sequences may be chemically synthesised and ligated together to provide a longer sequence.

In addition to being used as primers and/or probes, hybridising nucleic acid molecules for use in the present invention can be used as anti-sense molecules to alter the expression of polypeptides for use in the invention by binding to complementary nucleic acid molecules. This technique can be used in anti-sense therapy.

As used herein, an "antisense" nucleic acid refers to a nucleic acid capable of hybridising by virtue of some sequence complementarity to a portion of an RNA (preferably mRNA) encoding a polypeptide as defined herein. The antisense nucleic acid may be complementary to a coding and/or non-coding region of a mRNA encoding such a polypeptide. Such antisense nucleic acids have utility as compounds that inhibit expression, and can be used in the treatment or prevention of breast cancer and/or pancreatic cancer.

In a specific embodiment, expression of a polypeptide as defined herein is inhibited by use of antisense nucleic acids. The present invention provides the therapeutic or prophylactic use of nucleic acids comprising at least six nucleotides that are antisense to a gene or cDNA encoding NKCC1, a derivative or fragment thereof.

A hybridising nucleic acid molecule of use in the present invention may have a high degree of sequence identity along its length with a nucleic acid molecule within the scope of (a)-(e) above (e.g. at least 50%, at least 75%, at least 80%, at least 85%, at least 90% or at least 95% sequence identity). As will be appreciated by the skilled person, the higher the sequence identity a given single stranded nucleic acid molecule has with another nucleic acid molecule, the greater the likelihood that it will hybridise to a nucleic acid molecule which is complementary to that other nucleic acid molecule under appropriate conditions.

In view of the foregoing description the skilled person will appreciate that a large number of nucleic acids are within the scope of the present invention. Unless the context indicates otherwise, nucleic acid molecules for use in the present invention may have one or more of the following characteristics:

1) they may be DNA or RNA;
2) they may be single or double stranded;
3) they may be provided in recombinant form, e.g. covalently linked to a 5' and/or a 3' flanking sequence to provide a molecule which does not occur in nature;
4) they may be provided without 5' and/or 3' flanking sequences which normally occur in nature;
5) they may be provided in substantially pure form. Thus they may be provided in a form which is substantially free from contaminating proteins and/or from other nucleic acids; and
6) they may be provided with introns or without introns (e.g. as cDNA).

If desired, a gene encoding a polypeptide as defined herein, a related gene, or related nucleic acid sequences or subsequences, including complementary sequences, can also be used in hybridisation assays. A nucleotide encoding a polypeptide as defined herein, or subsequences thereof comprising at least 8 nucleotides, can be used as a hybridisation probe. Hybridisation assays can be used for detection, prognosis, diagnosis, or monitoring of conditions, disorders, or disease states, associated with aberrant expression of genes encoding a polypeptide as defined herein, or for differential diagnosis of patients with signs or symptoms suggestive of breast cancer and/or pancreatic cancer. In particular, such a hybridisation assay can be carried out by a method comprising contacting a patient sample containing nucleic acid with a nucleic acid probe capable

of hybridising to a DNA or RNA that encodes a polypeptide as defined herein, under conditions such that hybridisation can occur, and detecting or measuring any resulting hybridisation. Nucleotides can be used for therapy of patients having breast cancer and/or pancreatic cancer, as described below.

In another embodiment, a preparation of oligonucleotides comprising 10 or more consecutive nucleotides complementary to a nucleotide sequence encoding a polypeptide as defined herein or fragment thereof for use as vaccines for the treatment of breast cancer and/or pancreatic cancer. Such preparations may include adjuvants or other vehicles.

In a specific embodiment, nucleic acids comprising a sequence encoding a polypeptide as defined herein or functional derivative thereof, are administered to promote polypeptide function by way of gene therapy. Gene therapy refers to administration to a subject of an expressed or expressible nucleic acid. In this embodiment, the nucleic acid produces its encoded protein that mediates a therapeutic effect by promoting polypeptide function. Any of the methods for gene therapy available in the art can be used according to the present invention.

In a preferred aspect, the compound comprises a nucleic acid as defined herein, such as a nucleic acid encoding a polypeptide as defined herein or fragment or chimeric protein thereof, said nucleic acid being part of an expression vector that expresses a polypeptide as defined herein or fragment or chimeric protein thereof in a suitable host. In particular, such a nucleic acid has a promoter operably linked to the polypeptide coding region, said promoter being inducible or constitutive (and, optionally, tissue-specific). In another particular embodiment, a nucleic acid molecule is used in which the coding sequences and any other desired sequences are flanked by regions that promote homologous recombination at a desired site in the genome, thus providing for intrachromosomal expression of the nucleic acid (Koller & Smithies, 1989, *Proc. Natl. Acad. Sci. USA* 86:8932-8935; Zijlstra *et al.*, 1989, Nature 342:435-438).

Delivery of the nucleic acid into a patient may be direct, in which case the patient is directly exposed to the nucleic acid or nucleic acid-carrying vector; this approach is known as *in vivo* gene therapy. Alternatively, delivery of the nucleic acid into the patient may be indirect, in which case cells are first transformed with the nucleic acid *in vitro* and then transplanted into the patient; this approach is known as *ex vivo* gene therapy.

A convenient means for detecting/quantifying the polypeptides used in the present invention involves the use of antibodies. Thus, the polypeptides used in the invention also find use in raising antibodies. Thus, in a further aspect, the present invention provides the use of an antibody which binds to at least one polypeptide as defined in the first aspect of the invention for screening for and/or diagnosis of breast cancer and/or pancreatic cancer in a subject. Preferably, the antibody is used for detecting and/or quantifying the amount of a polypeptide as defined in the first aspect of the invention in a biological sample obtained from said subject.

In one embodiment, binding of antibody in tissue sections can be used to detect aberrant polypeptide localisation or an aberrant level of polypeptide. In a specific embodiment, an antibody to a polypeptide as defined herein can be used to assay a patient tissue (*e.g.*, a breast biopsy) for the level of the polypeptide where an aberrant level of polypeptide is indicative of breast cancer and/or pancreatic cancer. As used herein, an "aberrant level" means a level that is

increased or decreased compared with the level in a subject free from breast cancer and/or pancreatic cancer or a reference level.

Suitable immunoassays include, without limitation, competitive and non-competitive assay systems using techniques such as western blots, radioimmunoassays, ELISA (enzyme linked immunosorbent assay), "sandwich" immunoassays, immunoprecipitation assays, precipitin reactions, gel diffusion precipitin reactions, immunodiffusion assays, agglutination assays, complement-fixation assays, immunoradiometric assays, fluorescent immunoassays and protein A immunoassays.

In another aspect, the present invention provides a method for the prophylaxis and/or treatment of breast cancer and/or pancreatic cancer in a subject, which comprises administering to said subject a therapeutically effective amount of an antibody which binds to at least one polypeptide as defined in the first aspect of the invention.

In yet another aspect, the present invention provides the use of an antibody which binds to at least one polypeptide as defined in the first aspect of the invention in the preparation of a composition for use in the prophylaxis and/or treatment of breast cancer and/or pancreatic cancer. In particular, the preparation of vaccines and/or compositions comprising or consisting of antibodies is a preferred embodiment of this aspect of the invention.

Preferred antibodies bind specifically to polypeptides of the present invention so that they can be used to purify and/or inhibit the activity of such polypeptides. The antibodies may be monoclonal or polyclonal.

Thus, the polypeptide used in the invention, its fragments or other derivatives, or analogues thereof, may be used as an immunogen to generate antibodies which immunospecifically bind such an immunogen. Antibodies of the invention include, but are not limited to polyclonal, monoclonal, bispecific, humanised or chimeric antibodies, single chain antibodies, Fab fragments and F(ab') fragments, fragments produced by a Fab expression library, anti-idiotypic (anti-Id) antibodies, and epitope-binding fragments of any of the above. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically-active portions of immunoglobulin molecules, i.e., molecules that contain an antigen binding site that specifically binds an antigen. The immunoglobulin molecules of the invention can be of any class (e.g., IgG, IgE, IgM, IgD and IgA) or subclass of immunoglobulin molecule.

In the production of antibodies, screening for the desired antibody can be accomplished by techniques known in the art, e.g. ELISA (enzyme-linked immunosorbent assay). For example, to select antibodies, which recognise a specific domain of a polypeptide used in the invention, one may assay generated hybridomas for a product which binds to a polypeptide fragment containing such domain. For selection of an antibody that specifically binds a first polypeptide homologue but which does not specifically bind to (or binds less avidly to) a second polypeptide homologue, one can select on the basis of positive binding to the first polypeptide homologue and a lack of binding to (or reduced binding to) the second polypeptide homologue.

For preparation of monoclonal antibodies (mAbs) directed toward a polypeptide for use in the present invention, any technique which provides for the production of antibody molecules

by continuous cell lines in culture may be used. For example, the hybridoma technique originally developed by Kohler and Milstein (1975, Nature 256:495-497), as well as the trioma technique, the human B-cell hybridoma technique (Kozbor et al., 1983, Immunology Today 4:72), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole et al., 1985, in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96). Such antibodies may be of any immunoglobulin class including IgG, IgM, IgE, IgA, IgD and any subclass thereof. The hybridoma producing the mAbs in the invention may be cultivated in vitro or in vivo. In an additional embodiment of the invention, monoclonal antibodies can be produced in germ-free animals utilising known technology (PCT/US90/02545, incorporated herein by reference).

The mAbs include but are not limited to human mAbs and chimeric mAbs (e.g., human-mouse chimeras). A chimeric antibody is a molecule in which different portions are derived from different animal species, such as those having a human immunoglobulin constant region and a variable region derived from a murine mAb. (See, e.g., U.S. Patent No. 4,816,567; and U.S. Patent No. 4,816397, which are incorporated herein by reference in their entirety.) Humanised antibodies are antibody molecules from non-human species having one or more complementarity determining regions (CDRs) from the non-human species and a framework region from a human immunoglobulin molecule. (See, e.g., U.S. Patent No. 5,585,089, which is incorporated herein by reference in its entirety.)

Chimeric and humanized mAbs can be produced by recombinant DNA techniques known in the art, for example using methods described in WO 87/02671; EP 184,187; EP 171,496; EP 173,494;. WO 86/01533; U.S. 4,816,567; EP 125,023; Better et al., 1988, Science 240:1041-1043; Liu et al., 1987, Proc. Natl. Acad. Sci. USA 84:3439-3443; Liu et al., 1987, J. Immunol. 139:3521-3526; Sun et al., 1987, Proc. Natl. Acad. Sci. USA 84:214-218; Nishimura et al., 1987, Canc. Res. 47:999-1005; Wood et al., 1985, Nature 314:446-449; and Shaw et al., 1988, J. Natl. Cancer Inst. 80:1553-1559; Morrison, 1985, Science 229:1202-1207; Oi et al., 1986, Bio/Techniques 4:214; U.S. Patent 5,225,539; Jones et al., 1986, Nature 321:552-525; Verhoeyan et al. (1988) Science 239:1534; and Beidler et al., 1988, J. Immunol. 141:4053-4060.

Completely human antibodies are particularly desirable for therapeutic treatment of human patients. Such antibodies can be produced using transgenic mice which are incapable of expressing endogenous immunoglobulin heavy and light chain genes, but which can express human heavy and light chain genes. The transgenic mice are immunised in the normal fashion with a selected antigen, e.g., all or a portion of a polypeptide for use in the invention. mAbs directed against the antigen can be obtained using conventional hybridoma technology. The human immunoglobulin transgenes harboured by the transgenic mice rearrange during B cell differentiation, and subsequently undergo class switching and somatic mutation. Thus, using such a technique, it is possible to produce therapeutically useful IgG, IgA, IgM, IgD and IgE antibodies. For an overview of this technology for producing human antibodies, see Lonberg and Huszar (1995, Int. Rev. Immunol. 13:65-93). For a detailed discussion of this technology for producing human antibodies and human monoclonal antibodies and protocols for producing such antibodies, see, e.g., U.S. Patent 5,625,126; U.S. Patent 5,633,425; U.S. Patent 5,569,825; U.S. Patent 5,661,016; and U.S. Patent 5,545,806. In addition, companies such as Abgenix, Inc.

(Freemont, CA) and Genpharm (San Jose, CA) can be engaged to provide human antibodies directed against a selected antigen using technology similar to that described above.

Completely human antibodies, which recognise a selected epitope, can be generated using a technique referred to as "guided selection." In this approach a selected non-human monoclonal antibody, e.g., a mouse antibody, is used to guide the selection of a completely human antibody recognising the same epitope. (Jespers et al. (1994) Bio/technology 12:899-903).

The antibodies in the present invention can also be generated using various phage display methods known in the art. In phage display methods, functional antibody domains are displayed on the surface of phage particles which carry the polynucleotide sequences encoding them. In a particular, such phage can be utilised to display antigen binding domains expressed from a repertoire or combinatorial antibody library (e.g., human or murine). Phage expressing an antigen binding domain that binds the antigen of interest can be selected or identified with antigen, e.g., using labelled antigen or antigen bound or captured to a solid surface or bead. Phage used in these methods are typically filamentous phage including fd and M13 binding domains expressed from phage with Fab, Fv or disulphide stabilised Fv antibody domains recombinantly fused to either the phage gene III or gene VIII protein. Phage display methods that can be used to make the antibodies of the present invention include those disclosed in Brinkman et al., J. Immunol. Methods 182: 41-50 (1995); Ames et al., J. Immunol. Methods 184:177-186 (1995); Kettleborough et al., Eur. J. Immunol. 24:952-958 (1994); Persic et al., Gene 187 9-18 (1997); Burton et al., Advances in Immunology 57:191-280 (1994);. PCT/GB91/01134; WO 90/02809; WO 91/10737; WO 92/01047; WO 92/18619; WO 93/11236; WO 95/15982; WO 95/20401; and U.S. Patent Nos. 5,698,426; 5,223,409; 5,403,484; 5,580,717; 5,427,908; 5,750,753; 5,821,047; 5,571,698; 5,427,908; 5,516,637; 5,780,225; 5,658,727; 5,733,743 and 5,969,108; each of which is incorporated herein by reference in its entirety.

As described in the above references, after phage selection, the antibody coding regions from the phage can be isolated and used to generate whole antibodies, including human antibodies, or any other desired antigen binding fragment, and expressed in any desired host, including mammalian cells, insect cells, plant cells, yeast, and bacteria, e.g., as described in detail below. For example, techniques to recombinantly produce Fab, Fab' and F(ab')2 fragments can also be employed using methods known in the art such as those disclosed in WO 92/22324; Mullinax et al., BioTechniques 12(6):864-869 (1992); and Sawai et al., AJRI 34:26-34 (1995); and Better et al., Science 240:1041-1043 (1988) (said references incorporated by reference in their entireties).

Examples of techniques which can be used to produce single-chain Fvs and antibodies include those described in U.S. Patents 4,946,778 and 5,258,498; Huston et al., Methods in Enzymology 203:46-88 (1991); Shu et al., PNAS 90:7995-7999 (1993); and Skerra et al., Science 240:1038-1040 (1988).

The invention further provides for the use of bispecific antibodies, which can be made by methods known in the art. Traditional production of full length bispecific antibodies is based on the coexpression of two immunoglobulin heavy chain-light chain pairs, where the two chains have different specificities (Milstein et al., 1983, Nature 305:537-539). Because of the random

assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of 10 different antibody molecules, of which only one has the correct bispecific structure. Purification of the correct molecule, which is usually done by affinity chromatography steps, is rather cumbersome, and the product yields are low. Similar procedures are disclosed in WO 93/08829, and in Traunecker et al., 1991, EMBO J. 10:3655-3659.

According to a different and more preferred approach, antibody variable domains with the desired binding specificities (antibody-antigen combining sites) are fused to immunoglobulin constant domain sequences. The fusion preferably is with an immunoglobulin heavy chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light chain binding, present in at least one of the fusions. DNAs encoding the immunoglobulin heavy chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. This provides for great flexibility in adjusting the mutual proportions of the three polypeptide fragments in embodiments when unequal ratios of the three polypeptide chains used in the construction provide the optimum yields. It is, however, possible to insert the coding sequences for two or all three polypeptide chains in one expression vector when the expression of at least two polypeptide chains in equal ratios results in high yields or when the ratios are of no particular significance.

In a preferred embodiment of this approach, the bispecific antibodies are composed of a hybrid immunoglobulin heavy chain with a first binding specificity in one arm, and a hybrid immunoglobulin heavy chain-light chain pair (providing a second binding specificity) in the other arm. It was found that this asymmetric structure facilitates the separation of the desired bispecific compound from unwanted immunoglobulin chain combinations, as the presence of an immunoglobulin light chain in only one half of the bispecific molecule provides for a facile way of separation. This approach is disclosed in WO 94/04690. For further details for generating bispecific antibodies see, for example, Suresh et al., Methods in Enzymology, 1986, 121:210.

The invention provides functionally-active fragments, derivatives or analogues of the anti-polypeptide immunoglobulin molecules. "Functionally-active" means that the fragment, derivative or analogue is able to elicit anti-anti-idiotype antibodies (i.e., tertiary antibodies) that recognise the same antigen that is recognised by the antibody from which the fragment, derivative or analogue is derived. Specifically, in a preferred embodiment, the antigenicity of the idiotype of the immunoglobulin molecule may be enhanced by deletion of framework and CDR sequences that are C-terminal to the CDR sequence that specifically recognises the antigen. To determine which CDR sequences bind the antigen, synthetic peptides containing the CDR sequences can be used in binding assays with the antigen by any binding assay method known in the art.

The present invention provides antibody fragments such as, but not limited to, F(ab')2 fragments and Fab fragments. Antibody fragments which recognise specific epitopes may be generated by known techniques. F(ab')2 fragments consist of the variable region, the light chain constant region and the CH1 domain of the heavy chain and are generated by pepsin digestion of the antibody molecule. Fab fragments are generated by reducing the disulphide bridges of the F(ab')2 fragments. The invention also provides heavy chain and light chain dimers of the

antibodies of the invention, or any minimal fragment thereof such as Fvs or single chain antibodies (SCAs) (e.g., as described in U.S. Patent 4,946,778; Bird, 1988, Science 242:423-42; Huston et al., 1988, Proc. Natl. Acad. Sci. USA 85:5879-5883; and Ward et al., 1989, Nature 334:544-54), or any other molecule with the same specificity as the antibody of the invention. Single chain antibodies are formed by linking the heavy and light chain fragments of the Fv region via an amino acid bridge, resulting in a single chain polypeptide. Techniques for the assembly of functional Fv fragments in *E. coli* may be used (Skerra et al., 1988, Science 242:1038-1041).

In other embodiments, the invention provides fusion proteins of the immunoglobulins of the invention (or functionally active fragments thereof), for example in which the immunoglobulin is fused via a covalent bond (e.g., a peptide bond), at either the N-terminus or the C-terminus to an amino acid sequence of another protein (or portion thereof, preferably at least 10, 20 or 50 amino acid portion of the protein) that is not the immunoglobulin. Preferably the immunoglobulin, or fragment thereof, is covalently linked to the other protein at the N-terminus of the constant domain. As stated above, such fusion proteins may facilitate purification, increase half-life in vivo, and enhance the delivery of an antigen across an epithelial barrier to the immune system.

The immunoglobulins of the invention include analogues and derivatives that are either modified, i.e., by the covalent attachment of any type of molecule as long as such covalent attachment that does not impair immunospecific binding. For example, but not by way of limitation, the derivatives and analogues of the immunoglobulins include those that have been further modified, e.g., by glycosylation, acetylation, pegylation, phosphylation, amidation, derivatisation by known protecting/blocking groups, proteolytic cleavage, linkage to a cellular ligand or other protein, etc. Any of numerous chemical modifications may be carried out by known techniques, including, but not limited to specific chemical cleavage, acetylation, formylation, etc. Additionally, the analogue or derivative may contain one or more non-classical amino acids.

The foregoing antibodies can be used in methods known in the art relating to the localisation and activity of the polypeptides used in the invention, e.g., for imaging or radioimaging these proteins, measuring levels thereof in appropriate physiological samples, in diagnostic methods, etc. and for radiotherapy.

The antibodies of the invention can be produced by any method known in the art for the synthesis of antibodies, in particular, by chemical synthesis or by recombinant expression, and are preferably produced by recombinant expression technique.

Recombinant expression of antibodies, or fragments, derivatives or analogues thereof, requires construction of a nucleic acid that encodes the antibody. If the nucleotide sequence of the antibody is known, a nucleic acid encoding the antibody may be assembled from chemically synthesised oligonucleotides (e.g., as described in Kutmeier et al., 1994, BioTechniques 17:242), which, briefly, involves the synthesis of overlapping oligonucleotides containing portions of the sequence encoding antibody, annealing and ligation of those oligonucleotides, and then amplification of the ligated oligonucleotides by PCR.

Alternatively, the nucleic acid encoding the antibody may be obtained by cloning the antibody. If a clone containing the nucleic acid encoding the particular antibody is not available, but the sequence of the antibody molecule is known, a nucleic acid encoding the antibody may be obtained from a suitable source (e.g., an antibody cDNA library, or cDNA library generated from any tissue or cells expressing the antibody) by PCR amplification using synthetic primers hybridisable to the 3' and 5' ends of the sequence or by cloning using an oligonucleotide probe specific for the particular gene sequence.

If an antibody molecule that specifically recognises a particular antigen is not available (or a source for a cDNA library for cloning a nucleic acid encoding such an antibody), antibodies specific for a particular antigen may be generated by any method known in the art, for example, by immunising an animal, such as a rabbit, to generate polyclonal antibodies or, more preferably, by generating monoclonal antibodies. Alternatively, a clone encoding at least the Fab portion of the antibody may be obtained by screening Fab expression libraries (e.g., as described in Huse et al., 1989, Science 246:1275-1281) for clones of Fab fragments that bind the specific antigen or by screening antibody libraries (See, e.g., Clackson et al., 1991, Nature 352:624; Hane et al., 1997 Proc. Natl. Acad. Sci. USA 94:4937).

Once a nucleic acid encoding at least the variable domain of the antibody molecule is obtained, it may be introduced into a vector containing the nucleotide sequence encoding the constant region of the antibody molecule (see, e.g., WO 86/05807; WO 89/01036; and U.S. Patent No. 5,122,464). Vectors containing the complete light or heavy chain for co-expression with the nucleic acid to allow the expression of a complete antibody molecule are also available. Then, the nucleic acid encoding the antibody can be used to introduce the nucleotide substitution(s) or deletion(s) necessary to substitute (or delete) the one or more variable region cysteine residues participating in an intrachain disulphide bond with an amino acid residue that does not contain a sulphydryl group. Such modifications can be carried out by any method known in the art for the introduction of specific mutations or deletions in a nucleotide sequence, for example, but not limited to, chemical mutagenesis, in vitro site directed mutagenesis (Hutchinson et al., 1978, J. Biol. Chem. 253:6551), PCR based methods, etc.

In addition, techniques developed for the production of "chimeric antibodies" (Morrison et al., 1984, Proc. Natl. Acad. Sci. 81:851-855; Neuberger et al., 1984, Nature 312:604-608; Takeda et al., 1985, Nature 314:452-454) by splicing genes from a mouse antibody molecule of appropriate antigen specificity together with genes from a human antibody molecule of appropriate biological activity can be used. As described supra, a chimeric antibody is a molecule in which different portions are derived from different animal species, such as those having a variable region derived from a murine mAb and a human antibody constant region, e.g., humanised antibodies.

Once a nucleic acid encoding an antibody molecule has been obtained, the vector for the production of the antibody molecule may be produced by recombinant DNA technology using techniques well known in the art. Thus, methods for preparing the polypeptides used in the invention by expressing nucleic acid containing the antibody molecule sequences are described herein. Methods which are well known to those skilled in the art can be used to construct

expression vectors containing an antibody molecule coding sequences and appropriate transcriptional and translational control signals. These methods include, for example, in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. See, for example, the techniques described in Sambrook et al. (1990, Molecular Cloning, A Laboratory Manual, 2d Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY) and Ausubel et al. (eds., 1998, Current Protocols in Molecular Biology, John Wiley & Sons, NY).

The expression vector is transferred to a host cell by conventional techniques and the transfected cells are then cultured by conventional techniques to produce an antibody of the invention.

The host cells used to express a recombinant antibody of the invention may be either bacterial cells such as *Escherichia coli*, or, preferably, eukaryotic cells, especially for the expression of whole recombinant antibody molecule. In particular, mammalian cells such as Chinese hamster ovary cells (CHO), in conjunction with a vector such as the major intermediate early gene promoter element from human cytomegalovirus is an effective expression system for antibodies (Foecking et al., 198, Gene 45:101; Cockett et al., 1990, BioTechnology 8:2).

A variety of host-expression vector systems may be utilised to express an antibody molecule of the invention. Such host-expression systems represent vehicles by which the coding sequences of interest may be produced and subsequently purified, but also represent cells which may, when transformed or transfected with the appropriate nucleotide coding sequences, express the antibody molecule of the invention in situ. These include but are not limited to microorganisms such as bacteria (e.g., *E. coli, B. subtilis*) transformed with recombinant bacteriophage DNA, plasmid DNA or cosmid DNA expression vectors containing antibody coding sequences; yeast (e.g., *Saccharomyces, Pichia*) transformed with recombinant yeast expression vectors containing antibody coding sequences; insect cell systems infected with recombinant virus expression vectors (e.g., *baculovirus*) containing the antibody coding sequences; plant cell systems infected with recombinant virus expression vectors (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or transformed with recombinant plasmid expression vectors (e.g., Ti plasmid) containing antibody coding sequences; or mammalian cell systems (e.g., COS, CHO, BHK, HEK 293, 3T3 cells) harbouring recombinant expression constructs containing promoters derived from the genome of mammalian cells (e.g., metallothionein promoter) or from mammalian viruses (e.g., the adenovirus late promoter; the vaccinia virus 7.5K promoter).

In bacterial systems, a number of expression vectors may be advantageously selected depending upon the use intended for the antibody molecule being expressed. For example, when a large quantity of such a protein is to be produced, for the generation of pharmaceutical compositions comprising an antibody molecule, vectors which direct the expression of high levels of fusion protein products that are readily purified may be desirable. Such vectors include, but are not limited, to the E. coli expression vector pUR278 (Ruther et al., 1983, EMBO J. 2:1791), in which the antibody coding sequence may be ligated individually into the vector in frame with the lac Z coding region so that a fusion protein is produced; pIN vectors (Inouye & Inouye, 1985, Nucleic Acids Res. 13:3101-3109; Van Heeke & Schuster, 1989, J. Biol. Chem. 24:5503-5509);

and the like. pGEX vectors may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption and binding to a matrix glutathione-agarose beads followed by elution in the presence of free glutathione. The pGEX vectors are designed to include thrombin or factor Xa protease cleavage sites so that the cloned target gene product can be released from the GST moiety.

In an insect system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes. The virus grows in *Spodoptera frugiperda* cells. The antibody coding sequence may be cloned individually into non-essential regions (for example, the polyhedrin gene) of the virus and placed under control of an AcNPV promoter (for example, the polyhedrin promoter). In mammalian host cells, a number of viral-based expression systems (e.g., an adenovirus expression system) may be utilised.

As discussed above, a host cell strain may be chosen which modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific fashion desired. Such modifications (e.g., glycosylation) and processing (e.g., cleavage) of protein products may be important for the activity of the protein.

For long-term, high-yield production of recombinant antibodies, stable expression is preferred. For example, cells lines that stably express an antibody of interest can be produced by transfecting the cells with an expression vector comprising the nucleotide sequence of the antibody and the nucleotide sequence of a selectable (e.g., neomycin or hygromycin), and selecting for expression of the selectable marker. Such engineered cell lines may be particularly useful in screening and evaluation of compounds that interact directly or indirectly with the antibody molecule.

The expression levels of the antibody molecule can be increased by vector amplification (for a review, see Bebbington and Hentschel, The use of vectors based on gene amplification for the expression of cloned genes in mammalian cells in DNA cloning, Vol.3. (Academic Press, New York, 1987)). When a marker in the vector system expressing antibody is amplifiable, increase in the level of inhibitor present in culture of host cell will increase the number of copies of the marker gene. Since the amplified region is associated with the antibody gene, production of the antibody will also increase (Crouse et al., 1983, Mol. Cell. Biol. 3:257).

The host cell may be co-transfected with two expression vectors for use within the invention, the first vector encoding a heavy chain derived polypeptide and the second vector encoding a light chain derived polypeptide. The two vectors may contain identical selectable markers which enable equal expression of heavy and light chain polypeptides. Alternatively, a single vector may be used which encodes both heavy and light chain polypeptides. In such situations, the light chain should be placed before the heavy chain to avoid an excess of toxic free heavy chain (Proudfoot, 1986, Nature 322:52; Kohler, 1980, Proc. Natl. Acad. Sci. USA 77:2197). The coding sequences for the heavy and light chains may comprise cDNA or genomic DNA.

Once the antibody molecule of the invention has been recombinantly expressed, it may be purified by any method known in the art for purification of an antibody molecule, for example, by

chromatography (e.g., ion exchange chromatography, affinity chromatography such as with protein A or specific antigen, and sizing column chromatography), centrifugation, differential solubility, or by any other standard technique for the purification of proteins.

Alternatively, any fusion protein may be readily purified by utilising an antibody specific for the fusion protein being expressed. For example, a system described by Janknecht *et al.* allows for the ready purification of non-denatured fusion proteins expressed in human cell lines (Janknecht *et al.*, 1991, Proc. Natl. Acad. Sci. USA 88:8972-897). In this system, the gene of interest is subcloned into a vaccinia recombination plasmid such that the open reading frame of the gene is translationally fused to an amino-terminal tag consisting of six histidine residues. The tag serves as a matrix binding domain for the fusion protein. Extracts from cells infected with recombinant vaccinia virus are loaded onto Ni^{2+} nitriloacetic acid-agarose columns and histidine-tagged proteins are selectively eluted with imidazole-containing buffers.

In a preferred embodiment, antibodies of the invention or fragments thereof are conjugated to a diagnostic or therapeutic moiety. The antibodies can be used for diagnosis or to determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, radioactive nuclides, positron emitting metals (for use in positron emission tomography), and nonradioactive paramagnetic metal ions. See generally U.S. Patent No. 4,741,900 for metal ions which can be conjugated to antibodies for use as diagnostics according to the present invention. Suitable enzymes include horseradish peroxidase, alkaline phosphatase, beta-galactosidase, or acetylcholinesterase; suitable prosthetic groups include streptavidin, avidin and biotin; suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride and phycoerythrin; suitable luminescent materials include luminol; suitable bioluminescent materials include luciferase, luciferin, and aequorin; and suitable radioactive nuclides include ^{125}I, ^{131}I, ^{111}In, and ^{99}Tc.

Antibodies of the invention or fragments thereof can be conjugated to a therapeutic agent or drug moiety to modify a given biological response. The therapeutic agent or drug moiety is not to be construed as limited to classical chemical therapeutic agents. For example, the drug moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for example, a toxin such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; a protein such as tumour necrosis factor, α-interferon, β-interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator, a thrombotic agent or an anti-angiogenic agent, e.g., angiostatin or endostatin; or, a biological response modifier such as a lymphokine, interleukin-1 (IL-1), interleukin-2 (IL-2), interleukin-6 (IL-6), granulocyte macrophage colony stimulating factor (GM-CSF), granulocyte colony stimulating factor (G-CSF), nerve growth factor (NGF) or other growth factor.

Techniques for conjugating such therapeutic moiety to antibodies are well known, see, e.g., Arnon et al., "Monoclonal Antibodies For Immunotargeting Of Drugs In Cancer Therapy", in Monoclonal Antibodies And Cancer Therapy, Reisfeld et al. (eds.), pp. 243-56 (Alan R. Liss, Inc. 1985); Hellstrom et al., "Antibodies For Drug Delivery", in Controlled Drug Delivery (2nd

Ed.), Robinson et al. (eds.), pp. 623-53 (Marcel Dekker, Inc. 1987); Thorpe, "Antibody Carriers Of Cytotoxic Agents In Cancer Therapy: A Review", in Monoclonal Antibodies '84: Biological And Clinical Applications, Pinchera et al. (eds.), pp. 475-506 (1985); "Analysis, Results, And Future Prospective Of The Therapeutic Use Of Radiolabeled Antibody In Cancer Therapy", in Monoclonal Antibodies For Cancer Detection And Therapy, Baldwin et al. (eds.), pp. 303-16 (Academic Press 1985), and Thorpe et al., "The Preparation And Cytotoxic Properties Of Antibody-Toxin Conjugates", Immunol. Rev., 62:119-58 (1982).

Alternatively, an antibody can be conjugated to a second antibody to form an antibody heteroconjugate as described in U.S. 4,676,980.

An antibody with or without a therapeutic moiety conjugated to it can be used as a therapeutic that is administered alone or in combination with cytotoxic factor(s) and/or cytokine(s).

A further aspect of the invention provides methods of screening for active agents that modulate (*e.g.*, upregulate or downregulate) a characteristic of, *e.g.*, the expression or the enzymatic or binding activity, of a polypeptide for use in the invention.

The invention provides methods for identifying active agents (*e.g.*, chemical compounds, proteins, or peptides) that bind to a polypeptide for use in the invention or have a modulatory effect (e.g. stimulatory, inhibitory, up-regulation, downregulation) effect on the expression or activity of a polypeptide for use in the invention. Examples of candidate agents, include, but are not limited to, nucleic acids (*e.g.*, DNA and RNA), carbohydrates, lipids, proteins, peptides, peptidomimetics, agonists, antagonists, small molecules and other drugs. Active agents can be obtained using any of the numerous suitable approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam, 1997, Anticancer Drug Des. 12:145; U.S. 5,738,996; and U.S. 5,807,683, each of which is incorporated herein in its entirety by reference).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt et al., 1993, Proc. Natl. Acad. Sci. USA 90:6909; Erb et al., 1994, Proc. Natl. Acad. Sci. USA 91:11422; Zuckermann et al., 1994, J. Med. Chem. 37:2678; Cho et al., 1993, Science 261:1303; Carrell et al., 1994, Angew. Chem. Int. Ed. Engl. 33:2059; Carell et al., 1994, Angew. Chem. Int. Ed. Engl. 33:2061; and Gallop et al., 1994, J. Med. Chem. 37:1233, each of which is incorporated herein in its entirety by reference.

Libraries of compounds may be presented, *e.g.*, presented in solution (*e.g.*, Houghten, 1992, Bio/Techniques 13:412-421), or on beads (Lam, 1991, Nature 354:82-84), chips (Fodor, 1993, Nature 364:555-556), bacteria (U.S. Patent No. 5,223,409), spores (U.S. Patent Nos. 5,571,698; 5,403,484; and 5,223,409), plasmids (Cull et al., 1992, Proc. Natl. Acad. Sci. USA 89:1865-1869) or phage (Scott and Smith, 1990, Science 249:386-390; Devlin, 1990, Science

249:404-406; Cwirla et al., 1990, Proc. Natl. Acad. Sci. USA 87:6378-6382; and Felici, 1991, J. Mol. Biol. 222:301-310), each of which is incorporated herein in its entirety by reference.

In one embodiment, agents that interact with (i.e., bind to) a polypeptide for use in the invention are identified in a cell-based assay system. In accordance with this embodiment, cells expressing polypeptide for use in the invention are contacted with a candidate agent or a control agent and the ability of the candidate agent to interact with the polypeptide is determined. If desired, this assay may be used to screen a plurality (*e.g.* a library) of candidate agents. The cell, for example, can be of prokaryotic origin (*e.g.*, *E. coli*) or eukaryotic origin (*e.g.*, yeast or mammalian). Further, the cells can express the polypeptide for use in the invention endogenously or be genetically engineered to express the polypeptide. In some embodiments, the polypeptide for use in the invention or the candidate agent is labelled, for example with a radioactive label (such as ^{32}P, ^{35}S or ^{125}I) or a fluorescent label (such as fluorescein isothiocyanate, rhodamine, phycoerythrin, phycocyanin, allophycocyanin, o-phthaldehyde or fluorescamine) to enable detection of an interaction between a polypeptide and a candidate agent. The ability of the candidate agent to interact directly or indirectly with the polypeptide for use within the invention can be determined by methods known to those of skill in the art. For example, the interaction between a candidate agent and a polypeptide can be determined by flow cytometry, a scintillation assay, immunoprecipitation or western blot analysis.

In another embodiment, agents that interact with (i.e., bind to) a polypeptide for use in the invention are identified in a cell-free assay system. In accordance with this embodiment, a native or recombinant polypeptide for use in the invention or fragment thereof is contacted with a candidate agent or a control agent and the ability of the candidate agent to interact with the polypeptide is determined. If desired, this assay may be used to screen a plurality (*e.g.* a library) of candidate agents. Preferably, the polypeptide is first immobilised, by, for example, contacting the polypeptide with an immobilized antibody which specifically recognizes and binds it, or by contacting a purified preparation of polypeptide with a surface designed to bind proteins. The polypeptide may be partially or completely purified (*e.g.*, partially or completely free of other polypeptides) or part of a cell lysate. Further, the polypeptide may be a fusion protein comprising the polypeptide for use in the invention or a biologically active portion thereof and a domain such as glutathionine-S-transferase. Alternatively, the polypeptide can be biotinylated using techniques well known to those of skill in the art (*e.g.*, biotinylation kit, Pierce Chemicals; Rockford, IL). The ability of the candidate agent to interact with the polypeptide can be can be determined by methods known to those of skill in the art.

In another embodiment, a cell-based assay system is used to identify active agents that bind to or modulate the activity of a protein, such as an enzyme, or a biologically active portion thereof, which is responsible for the production or degradation of the polypeptide for use in the invention or is responsible for the post-translational modification of the polypeptide. In a primary screen, a plurality (*e.g.*, a library) of agents are contacted with cells that naturally or recombinantly express: (i) a polypeptide of use in the invention; and (ii) a protein that is responsible for processing of the polypeptide in order to identify compounds that modulate the production, degradation, or post-translational modification of the polypeptide. If desired, active

agents identified in the primary screen can then be assayed in a secondary screen against cells naturally or recombinantly expressing the specific polypeptide of interest. The ability of the candidate agent to modulate the production, degradation or post-translational modification of a polypeptide can be determined by methods known to those of skill in the art, including without limitation, flow cytometry, a scintillation assay, immunoprecipitation and western blot analysis.

In another embodiment, agents that competitively interact with (i.e., bind to) a polypeptide for use in the invention are identified in a competitive binding assay. In accordance with this embodiment, cells expressing the polypeptide are contacted with a candidate agent and an agent known to interact with the polypeptide; the ability of the candidate agent to competitively interact with the polypeptide is then determined. Alternatively, agents that competitively interact with (i.e., bind to) a polypeptide are identified in a cell-free assay system by contacting the polypeptide with a candidate agent and an agent known to interact with the polypeptide. As stated above, the ability of the candidate agent to interact with a polypeptide for use in the invention can be determined by methods known to those of skill in the art. These assays, whether cell-based or cell-free, can be used to screen a plurality (*e.g.*, a library) of candidate agents.

In another embodiment, active agents that modulate (i.e., upregulate or downregulate) the expression of a polypeptide for use in the invention are identified by contacting cells (*e.g.*, cells of prokaryotic origin or eukaryotic origin) expressing the polypeptide with a candidate agent or a control agent (*e.g.*, phosphate buffered saline (PBS)) and determining the expression of the polypeptide or mRNA encoding the polypeptide. The level of expression of a selected polypeptide or mRNA encoding polypeptide in the presence of the candidate agent is compared to the level of expression of the polypeptide or mRNA encoding the polypeptide in the absence of the candidate agent (*e.g.*, in the presence of a control agent). The candidate agent can then be identified as a modulator of the expression of the polypeptide based on this comparison. For example, when expression of the polypeptide or mRNA encoding the polypeptide is significantly greater in the presence of the candidate agent than in its absence, the candidate agent is identified as a stimulator of expression of the polypeptide or mRNA encoding the polypeptide. Alternatively, when expression of the polypeptide or mRNA encoding the polypeptide is significantly less in the presence of the candidate agent than in its absence, the candidate agent is identified as an inhibitor of the expression of the polypeptide or mRNA encoding the polypeptide. The level of expression of a polypeptide for use in the invention, or the mRNA that encodes it, can be determined by methods known to those of skill in the art based on the present description. For example, mRNA expression can be assessed by Northern blot analysis or RT-PCR, and protein levels can be assessed by western blot analysis.

In another embodiment, active agents that modulate the activity of a polypeptide are identified by contacting a preparation containing the polypeptide, or cells (*e.g.*, prokaryotic or eukaryotic cells) expressing the polypeptide with a candidate agent or a control agent and determining the ability of the candidate agent to modulate (*e.g.*, stimulate or inhibit) the activity of polypeptide. The activity of a polypeptide can be assessed by detecting its effect on a "downstream effector" for example, but without limitation, induction of a cellular signal

transduction pathway of the polypeptide (*e.g.* , intracellular Ca^{2+}, diacylglycerol, IP3, etc.), detecting catalytic or enzymatic activity of the target on a suitable substrate, detecting the induction of a reporter gene (*e.g.*, a regulatory element that is responsive to a polypeptide for use in the invention and is operably linked to a nucleic acid encoding a detectable marker, *e.g.*, luciferase), or detecting a cellular response, for example, cellular differentiation, or cell proliferation as the case may be, based on the present description, techniques known to those of skill in the art can be used for measuring these activities (see, *e.g.*, U.S. Patent No. 5,401,639, which is incorporated in its entirety herein by reference). The candidate agent can then be identified as a modulator of the activity of a polypeptide for use in the invention by comparing the effects of the candidate agent to the control agent. Suitable control agents include phosphate buffered saline (PBS) and normal saline (NS).

In another embodiment, active agents that modulate (i.e., upregulate or downregulate) the expression, activity or both the expression and activity of a polypeptide for use in the invention are identified in an animal model. Examples of suitable animals include, but are not limited to, mice, rats, rabbits, monkeys, guinea pigs, dogs and cats. Preferably, the animal used represents a model of breast cancer and/or pancreatic cancer. In accordance with this embodiment, the candidate agent or a control agent is administered (*e.g.*, orally, rectally or parenterally such as intraperitoneally or intravenously) to a suitable animal and the effect on the expression, activity or both expression and activity of the polypeptide is determined. Changes in the expression of a polypeptide can be assessed by any suitable method described above, based on the present description.

In yet another embodiment, a polypeptide for use in the invention is used as a "bait protein" in a two-hybrid assay or three hybrid assay to identify other proteins that bind to or interact with the polypeptide (see, *e.g.*, U.S. 5,283,317; Zervos et al. (1993) Cell 72:223-232; Madura et al. (1993) J. Biol. Chem. 268:12046-12054; Bartel et al. (1993) Bio/Techniques 14:920-924; Iwabuchi et al. (1993) Oncogene 8:1693-1696; and WO 94/10300). As those skilled in the art will appreciate, such binding proteins are also likely to be involved in the propagation of signals by the polypeptides for use in the invention as, for example, upstream or downstream elements of a signalling pathway involving the polypeptides for use in the invention.

This invention further provides novel active agents identified by the above-described screening assays described above and uses thereof for treatments as described herein.

The invention also provides for treatment or prevention of cancer by administration of a therapeutic compound. Such compounds include but are not limited to: polypeptide as defined herein and analogues and derivatives (including fragments) thereof; antibodies thereto; nucleic acids encoding a polypeptide as defined herein, analogues, or derivatives; antisense nucleic acids to a gene encoding a polypeptide as defined herein, and agonists and antagonists of a gene encoding a polypeptide as defined herein or agonists and antagonists of a polypeptide as defined herein. An important feature of the present invention is the identification of a gene encoding a polypeptide as defined herein involved in breast cancer and/or pancreatic cancer. Breast cancer and/or pancreatic cancer can be treated or prevented by administration of a therapeutic compound

that modulates activity or expression of a polypeptide as defined herein in the tissue of cancer patients.

As used herein "active agent" refers to the polypeptides for use in the invention and nucleic acid molecules encoding the polypeptides, antibodies against the polypeptides and agents e.g. small molecules, which modulate the expression of the polypeptides for use in the invention.

As discussed herein, active agents of the invention find use in the treatment or prophylaxis of breast cancer and/or pancreatic cancer. Thus, in a further aspect, the present invention provides a pharmaceutical composition comprising at least one active agent, optionally together with one or more pharmaceutically acceptable excipients, carriers or diluents. In one embodiment, the pharmaceutical composition is for use as a vaccine and so any additional components will be acceptable for vaccine use. In addition, the skilled person will appreciate that one or more suitable adjuvants may be added to such vaccine preparations.

The composition will usually be supplied as part of a sterile, pharmaceutical composition that will normally include a pharmaceutically acceptable carrier. This pharmaceutical composition may be in any suitable form (depending upon the desired method of administering it to a patient).

It may be provided in unit dosage form, will generally be provided in a sealed container and may be provided as part of a kit. Such a kit would normally (although not necessarily) include instructions for use. It may include a plurality of said unit dosage forms.

The pharmaceutical composition may be adapted for administration by any appropriate route, for example by the oral (including buccal or sublingual), rectal, nasal, topical (including buccal, sublingual or transdermal), vaginal or parenteral (including subcutaneous, intramuscular, intravenous or intradermal) route. Such compositions may be prepared by any method known in the art of pharmacy, for example by admixing the active ingredient with the carrier(s) or excipient(s) under sterile conditions.

Pharmaceutical compositions adapted for oral administration may be presented as discrete units such as capsules or tablets; as powders or granules; as solutions, syrups or suspensions (in aqueous or non-aqueous liquids; or as edible foams or whips; or as emulsions).

Suitable excipients for tablets or hard gelatine capsules include lactose, maize starch or derivatives thereof, stearic acid or salts thereof.

Suitable excipients for use with soft gelatine capsules include for example vegetable oils, waxes, fats, semi-solid, or liquid polyols etc.

For the preparation of solutions and syrups, excipients which may be used include for example water, polyols and sugars. For the preparation of suspensions, oils (e.g. vegetable oils) may be used to provide oil-in-water or water in oil suspensions.

Pharmaceutical compositions adapted for transdermal administration may be presented as discrete patches intended to remain in intimate contact with the epidermis of the recipient for a prolonged period of time. For example, the active ingredient may be delivered from the patch by iontophoresis as generally described in *Pharmaceutical Research*, 3(6):318 (1986).

Pharmaceutical compositions adapted for topical administration may be formulated as ointments, creams, suspensions, lotions, powders, solutions, pastes, gels, sprays, aerosols or oils. For infections of the eye or other external tissues, for example mouth and skin, the compositions are

preferably applied as a topical ointment or cream. When formulated in an ointment, the active ingredient may be employed with either a paraffinic or a water-miscible ointment base. Alternatively, the active ingredient may be formulated in a cream with an oil-in-water cream base or a water-in-oil base. Pharmaceutical compositions adapted for topical administration to the eye include eye drops wherein the active ingredient is dissolved or suspended in a suitable carrier, especially an aqueous solvent. Pharmaceutical compositions adapted for topical administration in the mouth include lozenges, pastilles and mouth washes.

Pharmaceutical compositions adapted for rectal administration may be presented as suppositories or enemas.

Pharmaceutical compositions adapted for nasal administration wherein the carrier is a solid include a coarse powder having a particle size for example in the range 20 to 500 microns which is administered in the manner in which snuff is taken, i.e. by rapid inhalation through the nasal passage from a container of the powder held close up to the nose. Suitable compositions wherein the carrier is a liquid, for administration as a nasal spray or as nasal drops, include aqueous or oil solutions of the active ingredient.

Pharmaceutical compositions adapted for administration by inhalation include fine particle dusts or mists which may be generated by means of various types of metered dose pressurised aerosols, nebulisers or insufflators.

Pharmaceutical compositions adapted for vaginal administration may be presented as pessaries, tampons, creams, gels, pastes, foams or spray compositions.

Pharmaceutical compositions adapted for parenteral administration include aqueous and non-aqueous sterile injection solution which may contain anti-oxidants, buffers, bacteriostats and solutes which render the composition substantially isotonic with the blood of the intended recipient; and aqueous and non-aqueous sterile suspensions which may include suspending agents and thickening agents. Excipients which may be used for injectable solutions include water, alcohols, polyols, glycerine and vegetable oils, for example. The compositions may be presented in unit-dose or multi-dose containers, for example sealed ampoules and vials, and may be stored in a freeze-dried (lyophilised) condition requiring only the addition of the sterile liquid carried, for example water for injections, immediately prior to use. Extemporaneous injection solutions and suspensions may be prepared from sterile powders, granules and tablets.

The pharmaceutical compositions may contain preserving agents, solubilising agents, stabilising agents, wetting agents, emulsifiers, sweeteners, colourants, odourants, salts (polypeptides for use in the present invention may themselves be provided in the form of a pharmaceutically acceptable salt), buffers, coating agents or antioxidants. They may also contain therapeutically active agents in addition to the polypeptide for use in the present invention.

Dosages of the polypeptide, nucleic acid or antibody for use in the present invention can vary between wide limits, depending upon the disease or disorder to be treated, the age and condition of the individual to be treated, etc. and a physician will ultimately determine appropriate dosages to be used. This dosage may be repeated as often as appropriate. If side effects develop the amount and/or frequency of the dosage can be reduced, in accordance with normal clinical practice.

In view of the importance of NKCC1 in breast and/ or pancreatic cancer the following form additional aspects of the present invention:

i) a method for monitoring/assessing breast and/or pancreatic cancer treatment in a patient, which comprises the step of determining the presence or absence and/or quantifying at least one polypeptide, at least one nucleic acid molecule or at least one antibody for use in the invention in a biological sample obtained from said patient;

ii) a method for the identification of metastatic breast and/or pancreatic cancer cells in a biological sample obtained from a subject, which comprises the step of determining the presence or absence and/or quantifying at least one polypeptide, at least one nucleic acid molecule or at least one antibody for use in the invention.

(iii) methods of treating breast and/or pancreatic cancer, comprising administering to a patient a therapeutically effective amount of a compound that modulates (e.g., upregulates or downregulates) or complements the expression or the biological activity (or both) of a polypeptide as defined herein in patients having breast and/or pancreatic cancer, in order to (a) prevent the onset or development of breast and/or pancreatic cancer; (b) prevent the progression of breast and/or pancreatic cancer; or (c) ameliorate the symptoms of breast and/or pancreatic cancer.

In the context of the present invention, the biological sample can be obtained from any source, such as a serum sample or a tissue sample, e.g. breast or pancreatic tissue. When looking for evidence of metastasis, one would look at major sites of breast or pancreatic metastasis such as: lymph nodes, liver, lung and/or bone.

The invention also provides diagnostic kits, comprising an antibody against a polypeptide as defined herein. In addition, such a kit may optionally comprise one or more of the following: (1) instructions for using the antibody for diagnosis, prognosis, therapeutic monitoring or any combination of these applications; (2) a labelled binding partner to the antibody; (3) a solid phase (such as a reagent strip) upon which the antibody is immobilised; and (4) a label or insert indicating regulatory approval for diagnostic, prognostic or therapeutic use or any combination thereof. If no labelled binding partner to the antibody is provided, the anti-polypeptide antibody itself can be labelled with a detectable marker, eg., a chemiluminescent, enzymatic, fluorescent, or radioactive moiety.

The invention also provides a kit comprising a nucleic acid probe capable of hybridising to RNA encoding a polypeptide as defined herein. In a specific embodiment, a kit comprises in one or more containers a pair of primers (*e.g.*, each in the size range of 6-30 nucleotides, more preferably 10-30 nucleotides and still more preferably 10-20 nucleotides) that under appropriate reaction conditions can prime amplification of at least a portion of a nucleic acid encoding a polypeptide as defined herein, such as by polymerase chain reaction (see *e.g.*, Innis *et al.*, 1990, PCR Protocols, Academic Press, Inc., San Diego, CA), ligase chain reaction (see EP 320,308) use of Qβ replicase, cyclic probe reaction, or other methods known in the art.

The invention provides methods for identifying active agents, candidate compounds or test compounds that bind to a polypeptide as defined herein or have a stimulatory or inhibitory effect on the expression or activity of a polypeptide as defined herein.

The present invention also provides assays for use in drug discovery in order to identify or verify the efficacy of compounds for treatment or prevention of breast cancer and/or pancreatic cancer. Test compounds can be assayed for their ability to modulate levels of a polypeptide as defined herein in a subject having breast cancer and/or pancreatic cancer. Compounds able to modulate levels of a polypeptide as defined herein in a subject having breast cancer and/or pancreatic cancer towards levels found in subjects free from breast cancer and/or pancreatic cancer or to produce similar changes in experimental animal models of breast cancer and/or pancreatic cancer can be used as lead compounds for further drug discovery, or used therapeutically. Expression of a polypeptide as defined herein can be assayed by, for example, immunoassays, gel electrophoresis followed by visualisation, detection of activity, or any other method taught herein or known to those skilled in the art. Such assays can be used to screen candidate drugs, in clinical monitoring or in drug development, where abundance of a polypeptide as defined herein can serve as a surrogate marker for clinical disease.

This invention further provides novel active agents identified by the above-described screening assays and uses thereof for treatments as described herein. In addition, the invention also provides the use of an active agent, which interacts with, or modulates the activity of a polypeptide as defined herein in the manufacture of a composition for the treatment of breast cancer and/or pancreatic cancer.

In one embodiment, one or more active agents are administered alone or in combination with one or more additional treatments or therapeutic compounds for breast cancer and/or pancreatic cancer. Examples of such treatments include, surgery and radiation therapy. Examples of therapeutic compounds include but are not limited to cyclophosphamide (Cytoxan™); methotrexate (Methotrexate™); 5-fluorouracil (5-FU); paclitaxel (Taxol); docetaxel (Taxotere™); vincristine (Oncovin™); vinblastine (Velban™); vinorelbine (Navelbine™); doxorubicin (Adriamycin); tamoxifen (Nolvadex™); toremifene (Fareston™); megestrol acetate(Megace™); anastrozole (Arimidex™); goserelin (Zoladex™); anti-HER2 monoclonal antibody (Herceptin™); capecitabine (Xeloda™) and raloxifene hydrochloride (Evista™).

The compounds of the invention include but are not limited to any compound, *e.g.*, a small organic molecule, protein, peptide, antibody, nucleic acid, etc. that restores the profile towards normal.

In another embodiment, symptoms of breast cancer and/or pancreatic cancer may be ameliorated by decreasing the level or activity of a polypeptide as defined herein by using gene sequences encoding a polypeptide as defined herein in conjunction with well-known gene "knock-out," ribozyme or triple helix methods to decrease gene expression of the polypeptide. In this approach, ribozyme or triple helix molecules are used to modulate the activity, expression or synthesis of the gene, and thus to ameliorate the symptoms of breast cancer and/or pancreatic cancer. Such molecules may be designed to reduce or inhibit expression of a mutant or non-

mutant target gene. Techniques for the production and use of such molecules are well known to those of skill in the art.

Endogenous polypeptide expression can also be reduced by inactivating or "knocking out" the gene encoding the polypeptide, or the promoter of such a gene, using targeted homologous recombination (*e.g.*, see Smithies, *et al.*, 1985, Nature 317:230-234; Thomas & Capecchi, 1987, Cell 51:503-512; Thompson *et al.*, 1989, Cell 5:313-321; and Zijlstra *et al.*, 1989, Nature 342:435-438). For example, a mutant gene encoding a non-functional polypeptide (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous gene (either the coding regions or regulatory regions of the gene encoding the polypeptide) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express the target gene *in vivo*. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the target gene. This approach can be adapted for use in humans, provided the recombinant DNA constructs are directly administered or targeted to the required site *in vivo* using appropriate viral vectors.

Preferred features of each aspect of the invention are as for each of the other aspects *mutatis mutandis*. The prior art documents mentioned herein are incorporated to the fullest extent permitted by law.

The invention will now be described with reference to the following examples, which should not in any way be construed as limiting the scope of the present invention. The examples refer to the figures in which:

Figure 1: shows the nucleotide and amino acid sequences of NKCC1 (GenBank accession: U30246; SwissProt accession: P55011). The tandem spectra used to identify NKCC1 in breast and pancreatic cancer cell line membrane preparations are shown in bold, italicised, and underlined. Masses assigned to NKCC1 are shown in bold and italicised (see below).

Figure 2: shows the normal tissue distribution of NKCC1 mRNA. Levels of mRNA in normal tissues were quantified by real time RT-PCR. mRNA levels are expressed as the number of copies ng^{-1} cDNA.

Figure 3: shows the expression of NKCC1 mRNA in normal and breast cancer tissues. Levels of NKCC1 mRNA in donor matched normal and adjacent tumour tissues were measured by real time RT-PCR. mRNA levels are expressed as the number of copies ng^{-1} cDNA.

Figure 4: shows the expression of NKCC1 mRNA in normal pancreatic tissues, pancreatic cancer tissues, and normal and tumour-derived pancreatic cell lines. Levels of NKCC1 mRNA were measured by real time RT-PCR. mRNA levels are expressed as the number of copies ng^{-1} cDNA.

Figure 5: shows immunohistochemical analysis of NKCC1 protein expression in sections of breast ductal carcinoma tissues. Invasive ductal carcinoma tissue sections from two separate donors is shown (A and B); the top panel shows NKCC1 immunostaining, the bottom panel shows an adjacent hematoxylin and eosin stained section. Note the NKCC1 immunoreactivity on the plasma membrane of the carcinoma cells.

Example 1: Identification of NKCC1

NKCC1 was isolated from breast cancer (MDA MB 486/BT474) and pancreatic (HPAFII) cell lines.

Cells were grown to confluence in 15cm^2 cell culture dishes before fractionation. Before harvest and extraction, the cells (approx. 2x10e9 cells) were washed three times with PBS-CM. Cells were scraped from culture dishes in ice-cold PBS-CM (5ml per 2x10e8 cells) using a plastic cell lifter. The cells were then centrifuged at 1000 x g for 5 minutes at +4°C. The supernatant was removed and the cells were resuspended in 10mls of homogenisation buffer (250mM Sucrose in 10mM HEPES, 1mM EDTA 1mM Vanadate, 0.02% Azide) followed by centrifugation at 1000 x g for 5 minutes at +4°C. The supernatant was removed. The cell pellet was then resuspended in 5 x packed cell volume with homogenisation buffer plus protease inhibitors (Sigma). A ball bearing homogeniser (BBH) (8.002mm ball) was chilled and rinsed with homogenisation buffer. The cell suspension was taken up in a 2ml syringe and attached to one side of the BBH. Another syringe was attached to the other side of the BBH. The cell mixture was fed through the chamber up to five times. Breakage of the cells was monitored using a microscope and when the cells were sufficiently lysed the resulting mixture was centrifuged at 1000 x g for 5 minutes at +4°.

The resulting supernatant (PNS) was retained and 1ml of homogenisation buffer was added to the nuclear pellet followed by centrifugation at 1000x g for 5 minutes. The latter two fractions were pooled and centrifuged at 3000 x g for 10 minutes at +4°C. The 3000 x g supernatant was layered onto a 2ml 60% sucrose cushion in SW40 or SW60 tube and centrifuged at 100 000 x g for 45 minutes with slow acceleration and deceleration. The crude plasma membrane was evident as a discrete layer on top of the sucrose cushion. The upper layer was removed (cytosol) and the plasma membrane was collected using a pasteur pipette. The % sucrose of crude plasma membrane fraction was determined using a refractometer. The membrane preparation was diluted with HEPES buffer to reduce the sucrose content to below 15%. The crude plasma membrane preparation was layered on preformed 15 to 60% sucrose gradient in SW40 tube and spun at 100 000 x g for 17 hours with slow acceleration and deceleration.

The sucrose gradient was fractionated using the gradient unloader (speed 0.5, distance 2.5, fractions 35). The protein content of the fractions was measured and 10 micrograms of protein was run on a 4-20% acrylamide 1D gel (Novex) and subject to western blotting with antibodies to Transferrin Receptor, Oxidoreductase II and Calnexin.

Plasma membrane fractions that had transferrin immunoreactivity but no oxidoreductase II or calnexin immunoreactivity were identified. These sucrose fractions were pooled and diluted

at least four times with 10mM HEPES, 1mM EDTA 1mM Vanadate, 0.02% Azide. The diluted sucrose fraction was added to a SW40 or SW60 tube and centrifuged at 100 000 x g for 45 minutes with slow acceleration and deceleration. The supernatant was removed from the membrane pellet and the pellet washed three times with PBS-CM. The membrane pellet was solubulized in 2% SDS in 63mM TrisHCl, pH 7.4. A protein assay wasperformed followed by the addition of mercaptoethanol (2% final), glycerol (10%) and bromopheneol blue (0.0025% final). The extracted protein sample was finally solubilized in 1D lysis buffer to a final protein concentration of 1 microgram/microlitre and the proteins separated by 1D PAGE.

Mass Spectrometry

Proteins excised from the 1D gel were digested with trypsin and analysed by MALDI-TOF-MS (Voyager STR, Applied Biosystems) using a 337 nm wavelength laser for desorption and the reflectron mode of analysis. Selected masses for NKCC1 were further characterised by tandem mass spectrometry using a QTOF-MS equipped with a nanospray ion source, (Micromass UK Ltd.). Prior to MALDI analysis the samples were desalted and concentrated using C18 Zip Tips™ (Millipore). Samples for tandem mass spectrometry (MS) were purified using a nano LC system (LC Packings, Amsterdam, The Netherlands) incorporating C18 SPE material.

Using the SEQUEST search program (Eng et al., 1994, J. Am. Soc. Mass Spectrom. 5:976-989), uninterpreted tandem mass spectra of tryptic digest peptides were searched against a database of public domain proteins constructed of protein entries in the non-redundant database held by the National Centre for Biotechnology Information (NCBI). This database is accessible at http://www.ncbi.nlm.nih.gov/ and also constructed of Expressed Sequence Tags entries (http://www.ncbi.nlm.nih.gov/dbEST/index.html). As a result of sequence database searching tandem amino acid sequences were found to match a SwissProt accession number: P55011 (Bumetanide-sensitive sodium-(potassium)-chloride cotransporter 1), see Figure 1, matched tandem sequences are shown in bold.

Alternatively sequences were identified using peptide mass data derived from mass spectrometer analysis and the MOWSE database search procedure. Peptide mass information can provide a 'fingerprint' signature sufficiently discriminating to allow for the unique and rapid identification of unknown sample proteins, independent of other analytical methods such as protein sequence analysis. Practical experience has shown that sample proteins can be uniquely identified using as few as 3- 4 experimentally determined peptide masses when screened against a fragment database derived from over 50,000 proteins (D.J.C. Pappin, P. Hojrup and A.J. Bleasby 'Rapid Identification of Proteins by Peptide-Mass Fingerprinting'. Current Biology (1993), vol 3, 327-332. and http://www.hgmp.mrc.ac.uk/Bioinformatics/Webapp/mowse/). The version of the code used for a MOWSE database search had the following modifications: the size of the parent protein is not included in the calculation such that large proteins such as titin no longer bias the score; instead the theoretical frequency of a peptide of mass (x) is estimated using the mass (x) of the peptide and the mean mass of an amino acid whilst allowing for a probability of 0.2 for a missed internal cleavage (by trypsin) and 0.1 for the probability of the occurrence of a proteolytic cleavage site.

The score assigned to a match on a peptide mass (x) is the logarithm of the probability of finding such a match at random and is inversely proportional to the frequency of fragments of that mass.

The score is thus calculated using the linear regression formula found by plotting the score for a match on peptide mass (x) against the mass (x). Two mass matches to predicted trypsin fragments were identified in this manner for NKCC1, which also matched to P55011, see Figure 1, the sequences which were identified through mass matching are shown in italics and underlined. The accuracy of this method was 20 ppm.

Example 2: Expression of NKCC1 mRNA in human tissues

Real time quantitative RT-PCRwas used (Heid, C.A., Stevens, J., Livak, K.J. & Williams, P.M. Real time quantitative PCR. *Genome Res.* **6**, 986-994 (1996); Morrison, T.B., Weis, J.J. & Wittwer, C.T. Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification. *Biotechniques* **24**, 954-958 (1998)) to analyse the distribution of NKCC1 mRNA in normal human tissues (Fig 2), donor matched tumour and adjacent normal tissues from seven breast cancer patients, (Fig 3), normal and tumour-derived pancreatic cell lines and tissue specimens (Fig 4).

Quantification of NKCC1 mRNA by RT-PCR

Real-time RT-PCR was used to quantitatively measure NKCC1 expression in normal human tissue mRNAs (Clontech), donor matched tumour and adjacent normal tissues from breast cancer patients, normal and tumour-derived pancreatic cell lines and tissue specimens. Ethical approval for the normal and tumour tissue samples was obtained at surgery (University of Oxford, UK). The primers used for PCR were as follows:

sense, 5' cacctactacctgcgcaccttc 3', (SEQ ID NO: 3)
antisense, 5' gaccacagcatctctggttgga 3'. (SEQ ID NO: 4)

Reactions contained 5ng cDNA (prepared using Superscript first strand synthesis for RT-PCR kit (Life Technologies)), SYBR green sequence detection reagents (PE Biosystems), sense and antisense primers, and were assayed on an ABI7700 sequence detection system (PE Biosystems). The PCR conditions were 1 cycle at 50°C for 2 min, 1 cycle at 95°C for 10 min, and 40 cycles of 95°C for 15s, 65°C for 1min. The accumulation of PCR product was measured in real time as the increase in SYBR green fluorescence, and the data were analysed using the Sequence Detector program v1.6.3 (PE Biosystems). Standard curves relating initial template copy number to fluorescence and amplification cycle were generated using the amplified PCR product as a template, and were used to calculate NKCC1 copy number in each sample. Overall the distribution of NKCC1 mRNA was low in normal tissues, with the highest levels of mRNA expression seen in mammary, prostate, testis and brain tissues (Figure 1).

The expression of NKCC1 mRNA in clinical breast carcinoma tissues was compared with the matched adjacent normal tissue from 7 breast cancer patients (Figure 3). NKCC1

expression was increased in all of the tumour samples, relative to their matched control tissue, with six of the seven samples showing a greater than 6-fold increase in expression.

NKCC1 mRNA expression was also analysed in normal and tumour pancreatic tissues and cell lines (Figure 4). NKCC1 expression was found to be significantly higher in the pancreatic tumour tissue(mean = 69 copies ng^{-1} cDNA) and tumour-derived cell lines (mean = 1013 copies ng^{-1} cDNA) than in the corresponding normal pancreatic tissues and cell lines (mean = 16 and 5 copies per ng^{-1} cDNA respectively).

Example 3: Immunohistochemical Analysis of NKCC1 in Breast Cancer Tissues

To further illustrate the involvement of NKCC1 in breast cancer, immunohistochemistry with a specific anti-NKCC1 antibody was used to investigate NKCC1 protein expression in sections of multiple donor breast ductal carcinoma tissues.

NKCC1 Immunohistochemistry

Immunohistochemical analysis was carried out on formalin-fixed paraffin-embedded tissue microarrays containing 1mm sections of breast carcinoma tissue from 55 donors as well as 20 sections of various normal tissues (Clinomics Laboratories Inc., 165 Tor Court, Pittsfield, MA 01201). Slides were deparafinised by two 5 min washes in xylene then rehydrated through successive graded ethanol solutions and washed for 5 mins in PBS. Antigen retrieval was achieved by immersing the slides in 0.01M citrate buffer (pH 6) and microwaving for 10 mins at full power (950W). In addition, detection with the antibody was improved by protease treatment of the tissue with Autozyme (AbCam) for 10 min at room temperature. The tissue was blocked in 10% donkey serum/PBS for 1 hour before addition of 1.5 µg/ml primary polyclonal antibody (in 2.5% donkey serum/PBS). Following 3 washes in PBS the tissue sections were incubated with biotin-conjugated secondary antibodies (Biotin-SP-conjugated AffiniPure Donkey anti-guinea pig, Jackson ImmunoResearch) diluted at 1:200 (2.5 µg/ml in 2.5% donkey serum/PBS) for 1 hour. Slides were washed 3 times in PBS and the tissue incubated with Streptavidin-HRP (Jackson ImmunoResearch) diluted 1:100 (5 µg/ml in 2.5% donkey serum/PBS), followed by three 5 min washes in PBS. Antibody signal was detected using DAB substrate solution (Dako Ltd.) according to the manufacturers' instructions. An adjacent tissue array was counterstained for hematoxylin and eosin (Dako Ltd.) and images were captured by a digital camera attached to a light microscope.

Antibody generation.

The anti-NKCC1 polyclonal antibody was raised in guinea pigs immunised with 2 specific peptides (ISL Ltd., UK). Peptide sequences were chosen for synthesis based on plots of hydrophobicity, antigenicity, surface probability, and weak homology to other known protein family members. Peptides were synthesised using Fmoc chemistry with a cysteine residue added to the end of each to enable specific thiol-reactive coupling of Keyhole Limpet Haemocyanin prior to immunisation. The peptides used were; SKKPKGFFGYKC and

SGESEPAKGSEEAKGC. The antibodies were affinity purified using columns of the above immobilised peptides.

In the paired images (A + B) shown in Figure 5 the upper panel demonstrates NKCC1 protein immunostaining in the breast cancer tissue, the lower panel .shows a hematoxylin and eosin stained adjacent tissue section demonstrating the tissue histology (carcinoma cells have large blue nuclei).

It is clearly apparent from these Figures that NKCC1 is specifically and highly expressed in the ductal carcinoma cells of the breast cancer tissue (compare with adjacent breast tissue). In total 55 breast cancer donor tissues were examined for NKCC1 immunoreactivity. Of these 7 demonstrated very high NKCC1 staining, 30 exhibited high/moderate staining, 13 exhibited weak staining, and 5 showed no staining in the carcinoma cells.

Thus, NKCC1 shows a restricted pattern of expression in normal human tissues, and is elevated in breast cancer tissues and pancreatic cancer tissues and c cell lines, suggesting that this protein has potential as a specific therapeutic target for breast and/or pancreatic cancer.

The present invention is not to be limited in terms of the particular embodiments described in this application, which are intended as single illustrations of individual aspects of the invention. Functionally equivalent methods and apparatus within the scope of the invention, in addition to those enumerated herein, will be apparent to those skilled in the art from the foregoing description and accompanying drawings. Such modifications and variations are intended to fall within the scope of the appended claims.

When a reference is made herein to a method of treating or preventing a disease or condition using a particular agent or combination of agents, it is to be understood that such a reference is intended to include the use of that agent or combination of agents in the preparation of a medicament for the treatment or prevention of the disease or condition.

Preferred features of each aspect of the invention are as for each of the other aspects *mutatis mutandis*.

The contents of each reference, patent and patent application cited in this application is hereby incorporated by reference in its entirety.

CLAIMS

1. A method of screening for and/or diagnosis of breast cancer and/ or pancreatic cancer in a subject and/or monitoring the effectiveness of breast cancer and/ or pancreatic cancer therapy, which method comprises the step of detecting and/or quantifying in a biological sample obtained from said subject:

(i) a polypeptide which:

a) comprises or consists of the amino acid sequence shown in Figure 1 (SEQ ID NO:1);

b) is a derivative having one or more amino acid substitutions, modifications, deletions or insertions relative to the amino acid sequence shown in Figure 1 (SEQ ID NO:1); or

c) is a fragment of a polypeptide as defined in a) or b) above, which is at least ten amino acids long; and/or

(ii) a nucleic acid molecule which:

d) comprises or consists of the DNA sequence shown in Figure 1 (SEQ ID NO:2) or its RNA equivalent;

e) a sequence which codes for a derivative or fragment of an amino acid molecule as defined in (i)(b) (SEQ ID NO:2).

f) a sequence which is complementary to the sequences of d) or e);

g) a sequence which codes for the same polypeptide, as the sequences of d) or e); or

h) a sequence which shows substantial identity with any of those of d), e), f) or g).

2. An antibody, functionally-active fragment, derivative or analogue thereof, that specifically binds to one or more polypeptides as defined in claim 1(i).

3. The method according to claim 1, wherein the polypeptide is detected and/or quantified using an antibody that specifically binds to one or more polypeptides as defined in claim 1(i).

4. An antibody according to claim 2 or the method of claim 3; wherein the antibody is monoclonal, polyclonal, chimeric, bispecific, humanised or is conjugated to a therapeutic moiety, a second antibody or a fragment thereof, a cytotoxic agent or a cytokine.

5. A method for the prophylaxis and/or treatment of breast cancer and/or pancreatic cancer in a subject, which comprises administering to said subject a therapeutically effective amount of:

i) at least one polypeptide as defined in claim 1(i),

ii) at least one nucleic acid molecule as defined in claim 1(ii), or

iii) at least one antibody as defined in claim 2 or 4.

6. The use of:

i) at least one polypeptide as defined in claim 1(i),

ii) at least one nucleic acid molecule as defined in claim 1(ii), or

iii) at least one antibody as defined in claim 2 or 4.

in the preparation of a composition for use in the prophylaxis and/or treatment of breast cancer and/or pancreatic cancer.

7. i) at least one polypeptide as defined in claim 1(i),

ii) at least one nucleic acid molecule as defined in claim 1(ii), or

iii) at least one antibody as defined in claim 2 or 4,

for use in the prophylaxis and/or treatment of breast cancer and/or pancreatic cancer.

8. The use as claimed in claim 6; wherein the composition is a vaccine.

9. The method of claim 5 or the use of claim 6 or 7, where the nucleic acid inhibits the expression of the polypeptide as defined in claim 1(i).

10. A method of screening for agents that interact with one or more polypeptides as defined in claim 1(i), said method comprising:

(a) contacting said polypeptide with a candidate agent; and

(b) determining whether or not the candidate agent interacts with said polypeptide.

11. The method according to claim 10, wherein the determination of interaction between the candidate agent and the polypeptide comprises quantitatively detecting binding of the candidate agent and said polypeptide.

12. A method of screening for agents that modulate

i) the expression of an polypeptide as defined in claim 1(i),

ii) the expression of a nucleic acid molecule as defined in claim 1(ii), or

iii) an antibody as defined in claim 2,

said method comprising:

a) comparing the expression or activity of said polypeptide, the expression of said nucleic acid molecule or the expression of said antibody, in the presence of a candidate agent with the expression or activity of said polypeptide, the expression of said nucleic acid molecule or the expression of said antibody, in the absence of the candidate agent or in the presence of a control agent; and

b) determining whether the candidate agent causes the expression or activity of said polypeptide, the expression of said nucleic acid molecule or the expression of said antibody, to change.

13. The method of claim 12 wherein the expression or activity level of said polypeptide, the expression level of said nucleic acid molecule or the expression of said antibody is compared with a predetermined reference range.

14. The method of claims 12 or 13 wherein step (b) additionally comprises selecting an agent which modulates the expression or activity of said polypeptide, the expression of said nucleic acid molecule or the expression of said antibody, for further testing or therapeutic or prophylactic use as an anti-breast cancer and/or anti pancreatic cancer agent.

15. An active agent identified by the method of any one of claims 10 to 14 which causes the expression or activity of said polypeptide, the expression of said nucleic acid molecule or the expression of said antibody, to change.

16. The use of an active agent according to claim 15 in the manufacture of a composition for the prophylaxis and/or treatment of breast cancer and/or pancreatic cancer.

17. An active agent according to claim 15 for use in the prophylaxis and/or treatment of breast cancer and/or pancreatic cancer.

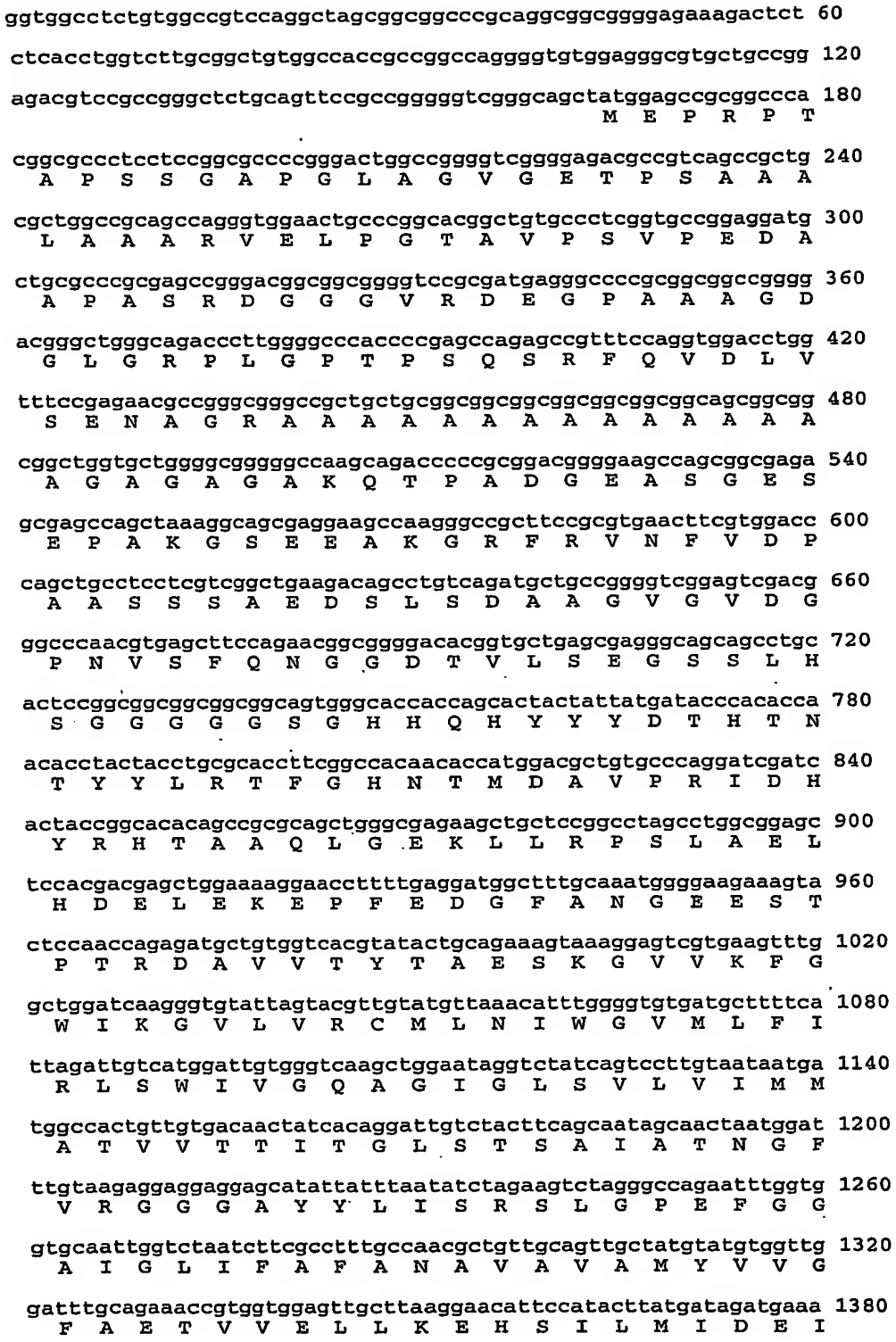
18. A method of prophylaxis and/or treatment of breast cancer and/or pancreatic cancer, which comprises administering to said subject a therapeutically effective amount of an active agent according to claim 15.

PROTEIN

Abstract

The present invention relates to the use of a protein (NKCC1) identified in breast cancer membrane preparations and pancreatic cancer membrane preparations, compositions comprising the protein, including vaccines and antibodies that are immunospecific for the protein. The use of the protein in the diagnosis, screening, treatment and prophylaxis of breast and/or pancreatic cancer is also provided.

Figure 1


ggtggcctctgtggccgtccaggctagcggcggcccgcaggcggcggggagaaagactct 60
ctcacctggtcttgcggctgtggccaccgccggccaggggtgtggagggcgtgctgccgg 120
agacgtccgccgggctctgcagttccgccgggggtcgggcagctatggagccgcggccca 180
M E P R P T
cggcgccctcctccggcgccccgggactggccggggtcggggagacgccgtcagccgctg 240
A P S S G A P G L A G V G E T P S A A A
cgctggccgcagccagggtggaactgcccggcacggctgtgccctcggtgccggaggatg 300
L A A A R V E L P G T A V P S V P E D A
ctgcgcccgcgagccgggacggcggcggggtccgcgatgagggccccgcggcggccgggg 360
A P A S R D G G G V R D E G P A A A G D
acgggctgggcagacccttggggcccaccccgagccagagccgtttccaggtggacctgg 420
G L G R P L G P T P S Q S R F Q V D L V
tttccgagaacgccgggcgggccgctgctgcggcggcggcggcggcggcggcagcggcgg 480
S E N A G R A A A A A A A A A A A A A A
cggctggtgctggggcgggggccaagcagacccccgcggacggggaagccagcggcgaga 540
A G A G A G A K Q T P A D G E A S G E S
gcgagccagctaaaggcagcgaggaagccaagggccgcttccgcgtgaacttcgtggacc 600
E P A K G S E E A K G R F R V N F V D P
cagctgcctcctcgtcggctgaagacagcctgtcagatgctgccggggtcggagtcgacg 660
A A S S S A E D S L S D A A G V G V D G
ggcccaacgtgagcttccagaacggcggggacacggtgctgagcgagggcagcagcctgc 720
P N V S F Q N G G D T V L S E G S S L H
actccggcggcggcggcggcagtgggcaccaccagcactactattatgatacccacacca 780
S G G G G G S G H H Q H Y Y Y D T H T N
acacctactacctgcgcaccttcggccacaacaccatggacgctgtgcccaggatcgatc 840
T Y Y L R T F G H N T M D A V P R I D H
actaccggcacacagccgcgcagctgggcgagaagctgctccggcctagcctggcggagc 900
Y R H T A A Q L G E K L L R P S L A E L
tccacgacgagctggaaaaggaacctttgaggatggctttgcaaatggggaagaaagta 960
H D E L E K E P F E D G F A N G E E S T
ctccaaccagagatgctgtggtcacgtatactgcagaaagtaaaggagtcgtgaagtttg 1020
P T R D A V V T Y T A E S K G V V K F G
gctggatcaagggtgtattagtacgttgtatgttaaacatttggggtgtgatgcttttca 1080
W I K G V L V R C M L N I W G V M L F I
ttagattgtcatggattgtgggtcaagctggaataggtctatcagtccttgtaataatga 1140
R L S W I V G Q A G I G L S V L V I M M
tggccactgttgtgacaactatcacaggattgtctacttcagcaatagcaactaatggat 1200
A T V V T T I T G L S T S A I A T N G F
ttgtaagaggaggaggagcatattatttaatatctagaagtctagggccagaatttggtg 1260
V R G G G A Y Y L I S R S L G P E F G G
gtgcaattggtctaatcttcgcctttgccaacgctgttgcagttgctatgtatgtggttg 1320
A I G L I F A F A N A V A V A M Y V V G
gatttgcagaaaccgtggtggagttgcttaaggaacattccatacttatgatagatgaaa 1380
F A E T V V E L L K E H S I L M I D E I

```
tcaatgatatccgaattattggagccattacagtcgtgattcttttaggtatctcagtag 1440
  N  D  I  R  I  I  G  A  I  T  V  V  I  L  L  G  I  S  V  A

ctggaatggagtgggaagcaaaagctcagattgttcttttggtgatcctacttcttgcta 1500
  G  M  E  W  E  A  K  A  Q  I  V  L  L  V  I  L  L  L  A  I

ttggtgatttcgtcataggaacatttatcccactggagagcaagaagccaaaagggtttt 1560
  G  D  F  V  I  G  T  F  I  P  L  E  S  K  K  P  K  G  F  F

ttggttataaatctgaaatatttaatgagaactttgggcccgattttcgagaggaagaga 1620
  G  Y  K  S  E  I  F  N  E  N  F  G  P  D  F  R  E  E  E  T

ctttcttttctgtatttgccatctttttttcctgctgcaactggtattctggctggagcaa 1680
  F  F  S  V  F  A  I  F  F  P  A  A  T  G  I  L  A  G  A  N

atatctcaggtgatcttgcagatcctcagtcagccatacccaaaggaacactcctagcca 1740
  I  S  G  D  L  A  D  P  Q  S  A  I  P  K  G  T  L  L  A  I

ttttaattactacattggtttacgtaggaattgcagtatctgtaggttcttgtgttgttc 1800
  L  I  T  T  L  V  Y  Y  V  G  I  A  V  S  V  G  S  C  V  V  R

gagatgccactggaaacgttaatgacactatcgtaacagagctaacaaactgtacttctg 1860
  D  A  T  G  N  V  N  D  T  I  V  T  E  L  T  N  C  T  S  A

cagcctgcaaattaaactttgatttttcatcttgtgaaagcagtccttgttcctatggcc 1920
  A  C  K  L  N  F  D  F  S  S  C  E  S  S  P  C  S  Y  G  L

taatgaacaacttccaggtaatgagtatggtgtcaggatttacaccactaatttctgcag 1980
  M  N  N  F  Q  V  M  S  M  V  S  G  F  T  P  L  I  S  A  G

gtatattttcagccactctttcttcagcattagcatccctagtgagtgctcccaaaatat 2040
  I  F  S  A  T  L  S  S  A  L  A  S  L  V  S  A  P  K  I  F

ttcaggctctatgtaaggacaacatctacccagctttccagatgtttgctaaaggttatg 2100
  Q  A  L  C  K  D  N  I  Y  P  A  F  Q  M  F  A  K  G  Y  G

ggaaaaataatgaacctcttcgtggctacatcttaacattcttaattgcacttggattca 2160
  K  N  N  E  P  L  R  G  Y  I  L  T  F  L  I  A  L  G  F  I

tcttaattgctgaactgaatgttattgcaccaattatctcaaacttcttccttgcatcat 2220
  L  I  A  E  L  N  V  I  A  P  I  I  S  N  F  F  L  A  S  Y

atgcattgatcaatttttcagtattccatgcatcacttgcaaaatctccaggatggcgtc 2280
  A  L  I  N  F  S  V  F  H  A  S  L  A  K  S  P  G  W  R  P

ctgcattcaaatactacaacatgtggatatcacttcttggagcaattctttgttgcatag 2340
  A  F  K  Y  Y  N  M  W  I  S  L  L  G  A  I  L  C  C  I  V

taatgttcgtcattaactggtgggctgcattgctaacatatgtgatagtccttgggctgt 2400
  M  F  V  I  N  W  W  A  A  L  L  T  Y  V  I  V  L  G  L  Y

atatttatgttacctacaaaaaaccagatgtgaattggggatcctctacacaagccctga 2460
  I  Y  V  T  Y  K  K  P  D  V  N  W  G  S  S  T  Q  A  L  T

cttacctgaatgcactgcagcattcaattcgtctttctggagtggaagaccacgtgaaaa 2520
  Y  L  N  A  L  Q  H  S  I  R  L  S  G  V  E  D  H  V  K  N

actttaggccacagtgtcttgttatgacaggtgctccaaactcacgtccagctttacttc 2580
  F  R  P  Q  C  L  V  M  T  G  A  P  N  S  R  P  A  L  L  H

atcttgttcatgatttcacaaaaaatgttggtttgatgatctgtggccatgtacatatgg 2640
  L  V  H  D  F  T  K  N  V  G  L  M  I  C  G  H  V  H  M  G

gtcctcgaagacaagccatgaaagagatgtccatcgatcaagccaaatatcagcgatggc 2700
  P  R  R  Q  A  M  K  E  M  S  I  D  Q  A  K  Y  Q  R  W  L

ttattaagaacaaaatgaaggcattttatgctccagtacatgcagatgacttgagagaag 2760
  I  K  N  K  M  K  A  F  Y  A  P  V  H  A  D  D  L  R  E  G

gtgcacagtatttgatgcaggctgctggtcttggtcgtatgaagccaaacacacttgtcc 2820
```

```
                A  Q  Y  L  M  Q  A  A  G  L  G  R  M  K  P  N  T  L  V  L

ttggatttaagaaagattggttgcaagcagatatgagggatgtggatatgtatataaact 2880
  G  F  K  K  D  W  L  Q  A  D  M  R  D  V  D  M  Y  I  N  L

tatttcatgatgcttttgacatacaatatggagtagtggttattcgcctaaaagaaggtc 2940
  F  H  D  A  F  D  I  Q  Y  G  V  V  V  I  R  L  K  E  G  L

tggatatatctcatcttcaaggacaagaagaattattgtcatcacaagagaaatctcctg 3000
  D  I  S  H  L  Q  G  Q  E  E  L  L  S  S  Q  E  K  S  P  G

gcaccaaggatgtggtagtaagtgtggaatatagtaaaaagtccgatttagatacttcca 3060
  T  K  D  V  V  V  S  V  E  Y  S  K  K  S  D  L  D  T  S  K

aaccactcagtgaaaaaccaattacacacaaagttgaggaagaggatggcaagactgcaa 3120
  P  L  S  E  K  P  I  T  H  K  V  E  E  E  D  G  K  T  A  T

ctcaaccactgttgaaaaaagaatccaaaggccctattgtgcctttaaatgtagctgacc 3180
  Q  P  L  L  K  K  E  S  K  G  P  I  V  P  L  N  V  A  D  Q

aaaagcttcttgaagctagtacacagtttcagaaaaaacaaggaaagaatactattgatg 3240
  K  L  L  E  A  S  T  Q  F  Q  K  K  Q  G  K  N  T  I  D  V

tctggtggctttttgatgatggaggtttgaccttattgataccttaccttctgacgacca 3300
  W  W  L  F  D  D  G  G  L  T  L  L  I  P  Y  L  L  T  T  K

agaaaaaatggaaagactgtaagatcagagtattcattggtggaaagataaacagaatag 3360
  K  K  W  K  D  C  K  I  R  V  F  I  G  G  K  I  N  R  I  D

accatgaccggagagcgatggctactttgcttagcaagttccggatagacttttctgata 3420
  H  D  R  R  A  M  A  T  L  L  S  K  F  R  I  D  F  S  D  I

tcatggttctaggagatatcaataccaaaccaaagaaagaaaatattatagcttttgagg 3480
  M  V  L  G  D  I  N  T  K  P  K  K  E  N  I  I  A  F  E  E

aaatcattgagccatacagacttcatgaagatgataaagagcaagatattgcagataaaa 3540
  I  I  E  P  Y  R  L  H  E  D  D  K  E  Q  D  I  A  D  K  M

tgaaagaagatgaaccatggcgaataacagataatgagcttgaactttataagaccaaga 3600
  K  E  D  E  P  W  R  I  T  D  N  E  L  E  L  Y  K  T  K  T

cataccggcagatcaggttaaatgagttattaaaggaacattcaagcacagctaatatta 3660
  Y  R  Q  I  R  L  N  E  L  L  K  E  H  S  S  T  A  N  I  I

ttgtcatgagtctcccagttgcacgaaaaggtgctgtgtctagtgctctctacatggcat 3720
  V  M  S  L  P  V  A  R  K  G  A  V  S  S  A  L  Y  M  A  W

ggttagaagctctatctaaggacctaccaccaatcctcctagttcgtgggaatcatcaga 3780
  L  E  A  L  S  K  D  L  P  P  I  L  L  V  R  G  N  H  Q  S

gtgtccttaccttctattcataaatgttctatacagtggacagccctccagaatggtact 3840
  V  L  T  F  Y  S  *

tcagtgcctagtgtagtaacctgaaatcttcaatgacacattaacatcacaatggcgaat 3900

ggtgacttttctttcacgatttcattaatttgaaagcacacaggaaagcttgctccattg 3960

ataacgtgtatggagacttcggttttagtcaattccatatctcaatcttaatggtgattc 4020

ttctctgttgaactgaagtttgtgagagtagttttcctttgctacttgaatagcaataaa 4080

agcgtgttaacttttttgg                                           4098
```

Figure 2

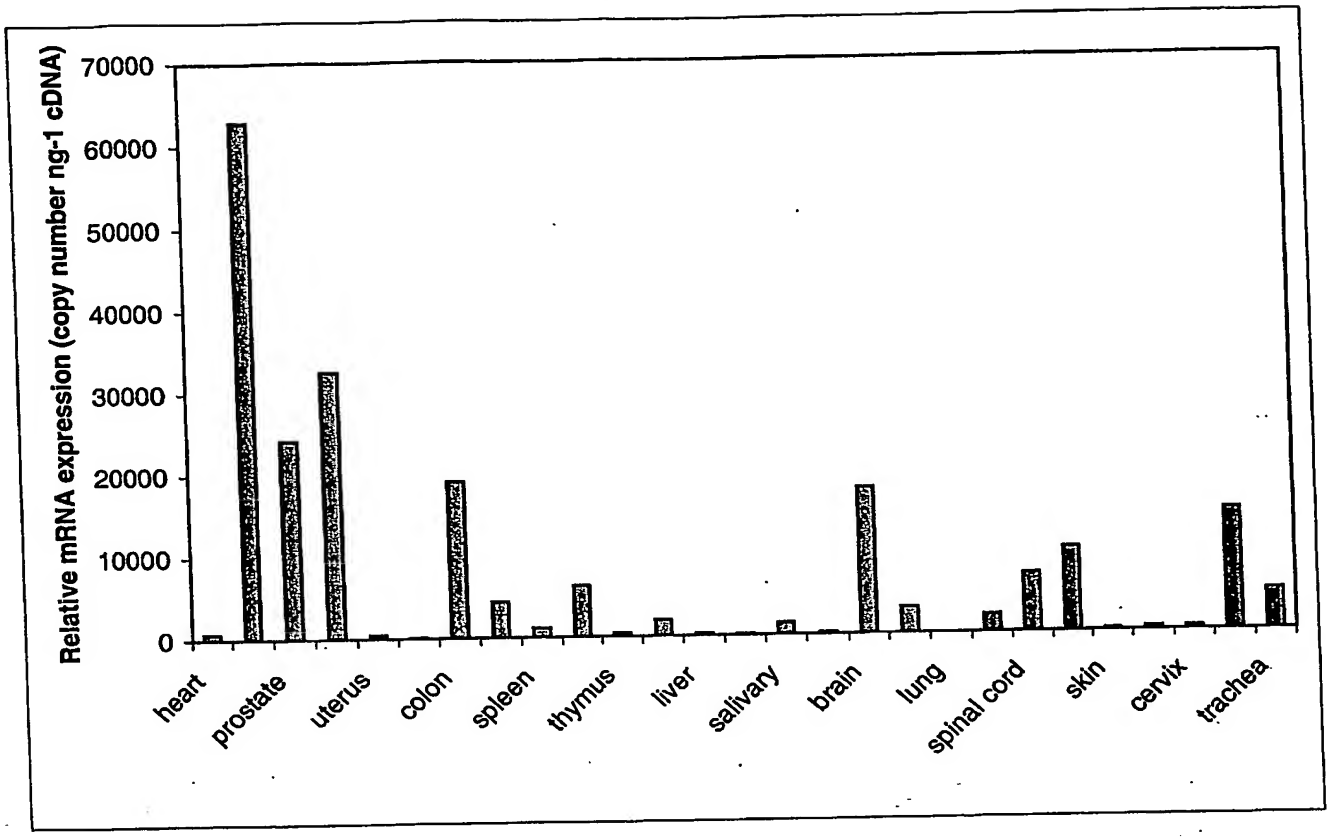

Relative mRNA expression (copy number ng-1 cDNA)
70000
60000
50000
40000
30000
20000
10000
0
heart
prostate
uterus
colon
spleen
thymus
liver
salivary
brain
lung
spinal cord
skin
cervix
trachea

Figure 3

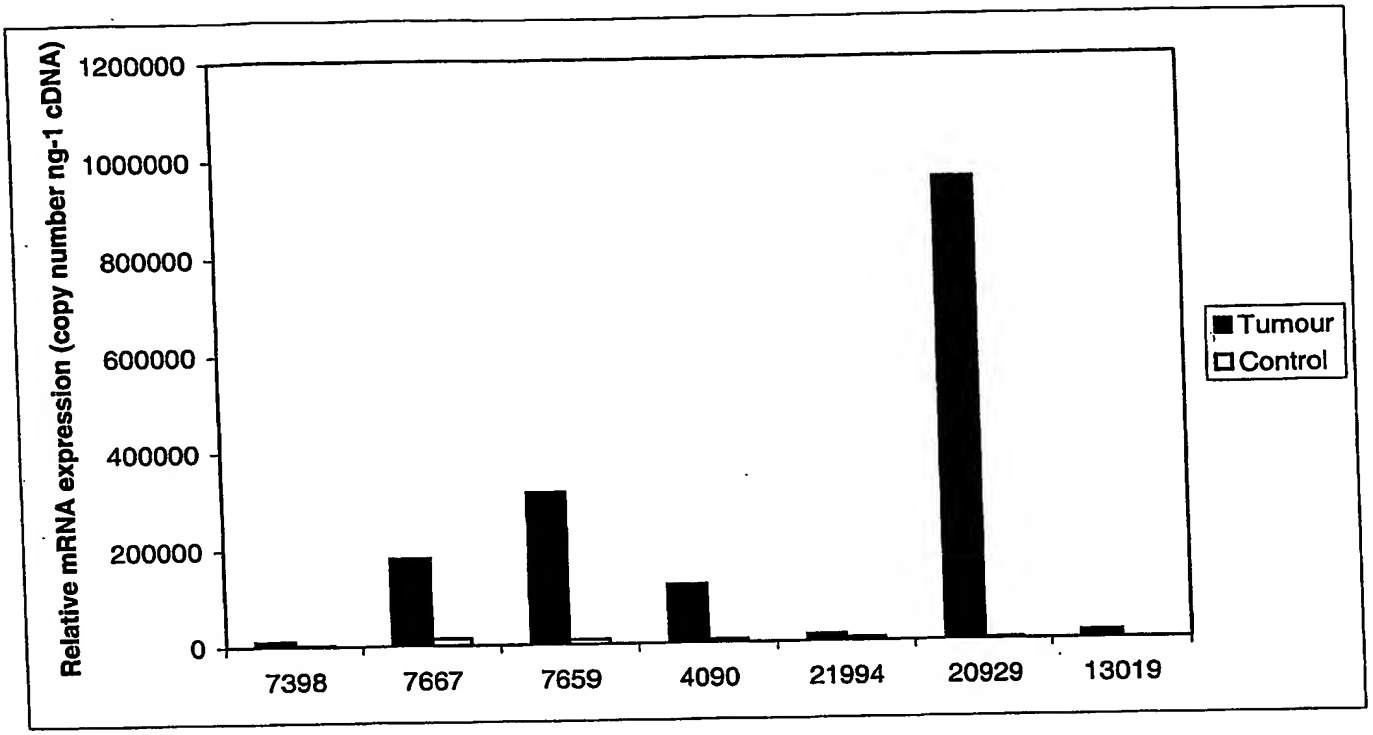

Relative mRNA expression (copy number ng-1 cDNA)
1200000
1000000
800000
600000
400000
200000
0
7398
7667
7659
4090
21994
20929
13019
Tumour
Control

Figure 4

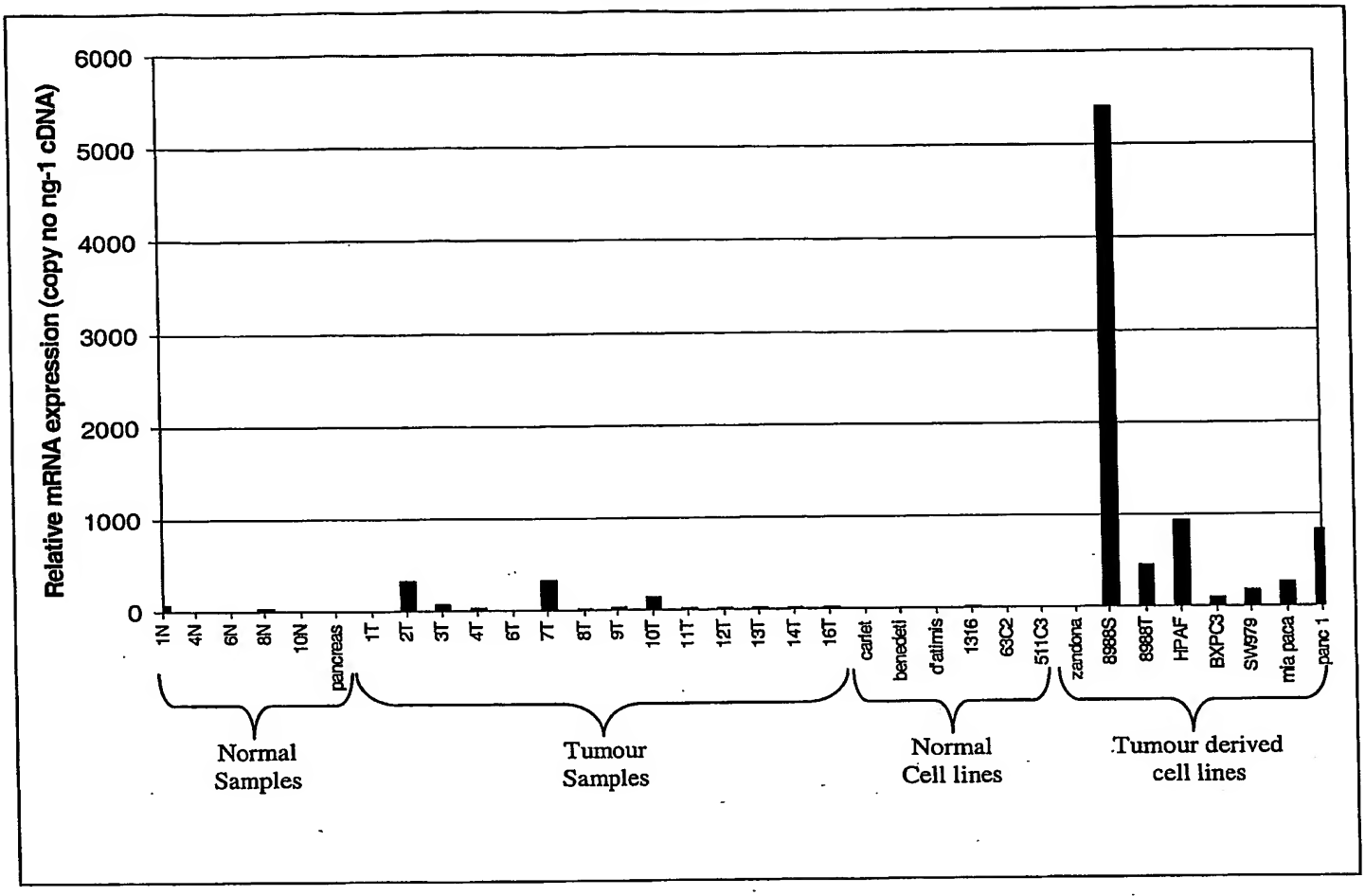
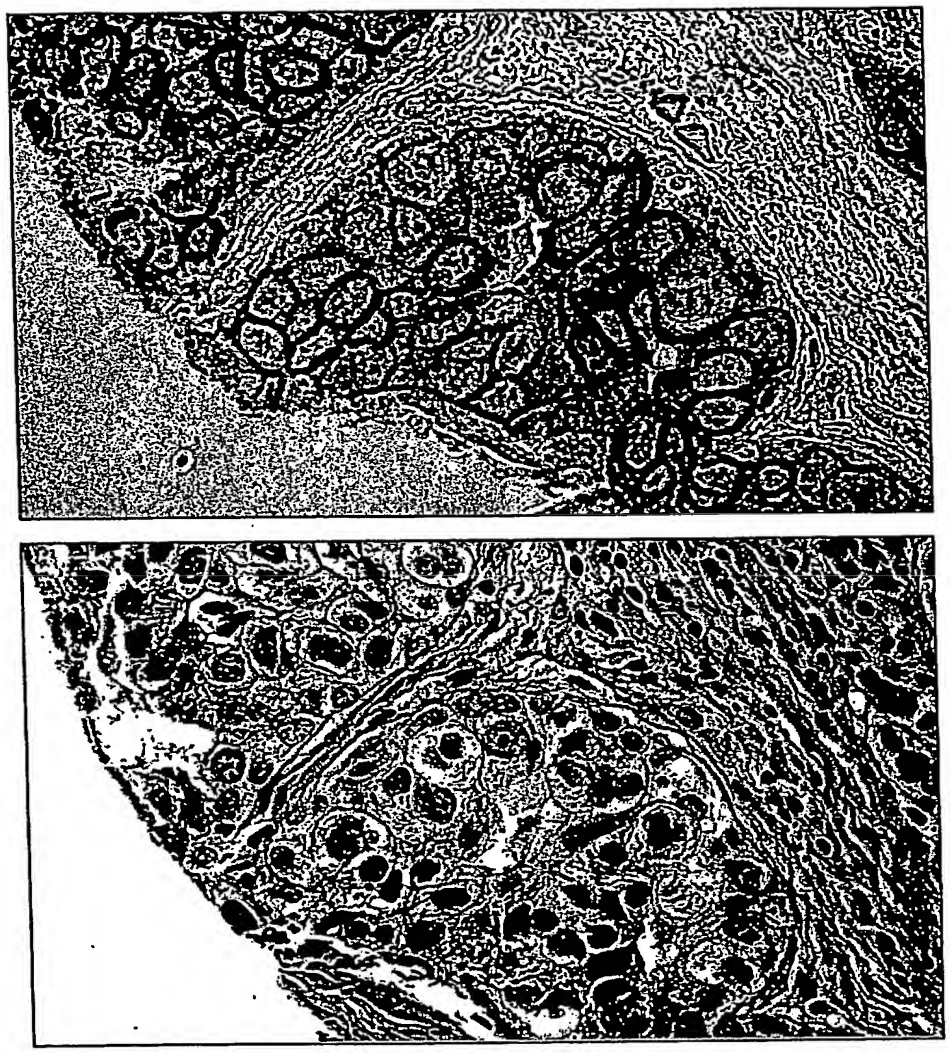
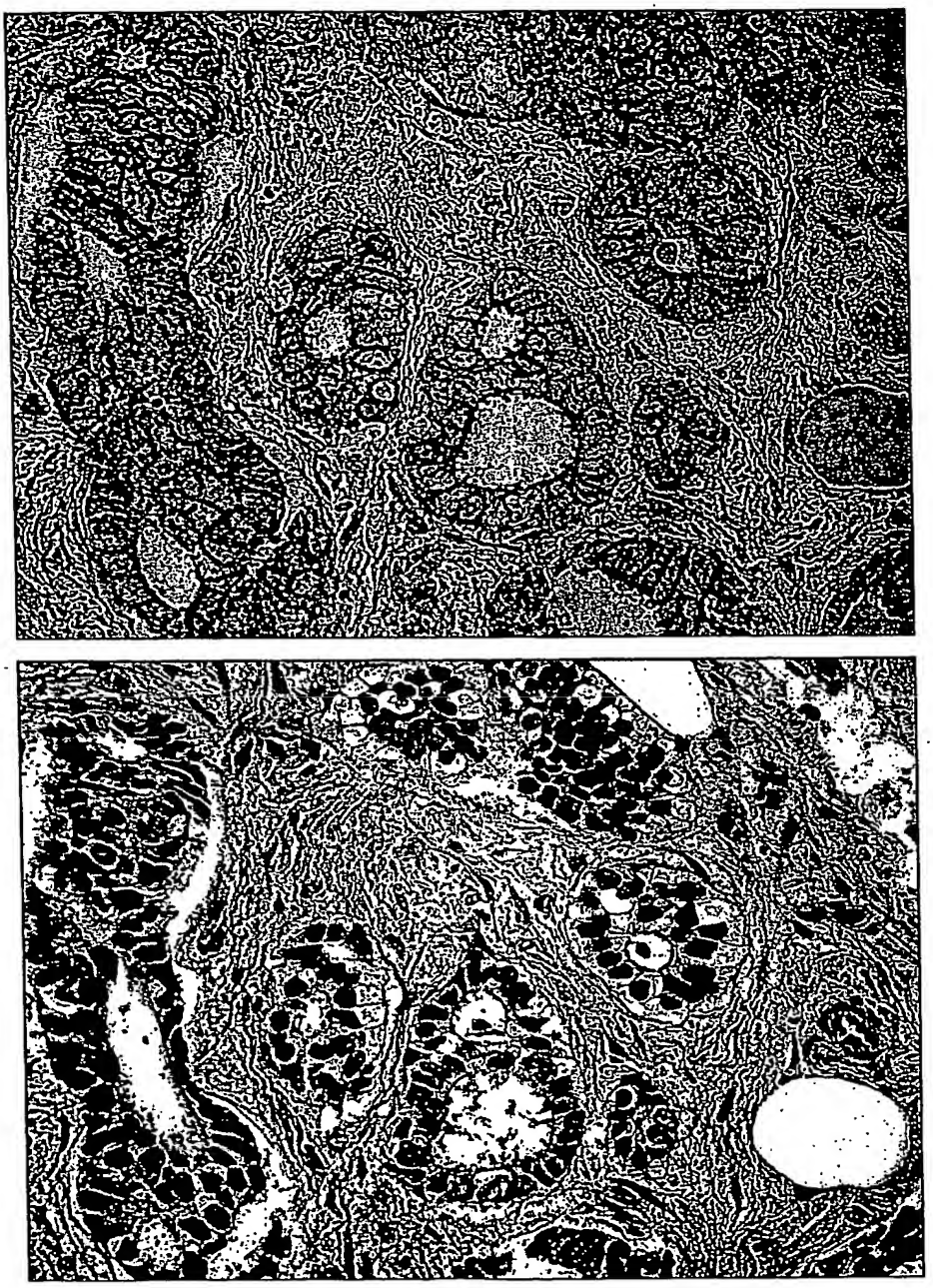

Relative mRNA expression (copy no ng-1 cDNA)
6000
5000
4000
3000
2000
1000
0
1N
4N
6N
8N
10N
pancreas
1T
2T
3T
4T
6T
7T
8T
9T
10T
11T
12T
13T
14T
16T
carlet
benedetl
d'atlmis
1316
63C2
511C3
zandona
8988S
8988T
HPAF
BXPC3
SW979
mia paca
panc 1
Normal Samples
Tumour Samples
Normal Cell lines
Tumour derived cell lines

Figure 5

A



B



This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☑ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.